**Universidade Federal de São Carlos**
**Centro de Ciências Exatas e de Tecnologia**
**Programa de Pós-Graduação em Ciência da Computação**

# Avaliação dos Mecanismos de Privacidade e Personalização na Web

Luanna Lopes Lobato

São Carlos
Maio/2007

**Universidade Federal de São Carlos**
**Centro de Ciências Exatas e de Tecnologia**
**Programa de Pós-Graduação em Ciência da Computação**

# Avaliação dos Mecanismos de Privacidade e Personalização na Web

Dissertação apresentada ao Programa de pós-graduação em Ciência da Computação do Departamento de Computação da Universidade Federal de São Carlos como parte dos requisitos para obtenção do título de Mestre.

Orientador: Dr. Sérgio Donizetti Zorzo

São Carlos
Maio/2007

# Universidade Federal de São Carlos

## Centro de Ciências Exatas e de Tecnologia

**Programa de Pós-Graduação em Ciência da Computação**

## *"Avaliação dos Mecanismos de Privacidade e Personalização na Web"*

**LUANNA LOPES LOBATO**

**Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.**

**Membros da Banca:**

**Prof. Dr. Sérgio Donizetti Zorzo**
**(Orientador – DC/UFSCar)**

**Prof. Dr. Hélio Crestana Guardia**
**(DC/UFSCar)**

**Profa. Dra. Lucila Ishitani**
**(PUC-MINAS)**

**São Carlos**
**Maio/2007**

**Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária da UFSCar**

L796am

Lobato, Luanna Lopes.
Avaliação dos mecanismos de Privacidade e Personalização na Web / Luanna Lopes Lobato. -- São Carlos : UFSCar, 2007.
146 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2007.

1. Privacidade e personalização. 2. Internet (Redes de computação). 3. Word wide web - medidas de segurança. I. Título.

CDD: 004.6 (20ª)

Aos meus pais e aos meus irmãos  pelo
exemplo de vida, de amor, de fé e pelo apoio.

... ao vento...
que me trouxe aqui. =]

“Se me acredito capaz de fazer algo, o farei, mesmo que eu não tenha capacidade ao começar...” (Mahatma Gandhi)

# Agradecimento

Em primeiro lugar agradeço a Deus por ter dado-me sabedoria, fé, persistência para que eu não desistisse de meus sonhos, capacidade para que eu pudesse realizá-los e por permitir que eu o sinta sempre presente em minha vida. 🕊

Aos meus pais, Luiz e Horacilda, meus maiores inspiradores, educadores e mestres, pela confiança, carinho, palavras de força, entendimento por eu ficar tanto tempo longe de casa, pelas orações atribuídas a mim e pelo imenso amor. Sem o incentivo deles eu jamais teria tido coragem e força para continuar minha caminhada. ♥

Agradeço aos meus irmãos por sempre acreditarem em mim e torcerem por meu sucesso. Ao meu maninho Zarthur (Luiz Arthur), o qual cuidou de mim durante os quatro anos de faculdade, me ensinando tantas coisas boas, as quais me deram base para conseguir ficar longe de casa. À minha maninha linda Ludy (Ludmylla), por ser tão especial, meiga, me dando sempre tanto carinho, fazendo-me sempre querer ser e fazer o melhor que eu possa, já que vejo em seus olhinhos uma imensa admiração por mim e um grande amor. ☼

Às minhas avós, Vó Donana (Ana) e Vó Luzia, por todo o amor que sempre me deram, por me receber com lágrimas nos olhos de tanta felicidade toda vez que eu chegava em casa, me fazendo ver o quanto sou amada por elas. Em especial, aos meus avôs, Vô Turzin (Artur) e Vô Tonhó (Antônio), que mesmo não estando mais presentes estão sempre em meu coração, em meus pensamentos e em meus sonhos, por terem cuidado de mim como uma filha. 🖋

Agradeço a todos meus amigos de Taparuba - MG, por me desejarem tão bem. ☺ Aos meus amigos Gau, Jenny, Marcos, Amanda, Dani e Leo amizade, carinho, respeito e amor que se fortificaram o bastante para nunca acabar. ♫ Agradeço ao Thiago pelo carinho, respeito e principalmente pelo companheirismo, me ajudando a tomar decisões importantes em minha vida, por ser uma pessoa tão doce, sincera e por sempre só me fazer o bem. ★

Aos professores do mestrado, pelo compartilhamento de conhecimento e em especial, ao meu orientador Zorzo pelos ensinamentos, pela atenção e por ter me indicado um ótimo tema de trabalho, sendo prazerosos meus estudos. Agradeço aos meus professores da graduação, Alexandre Konzen, André Gustavo e Ciro Menezes por terem sentido a minha felicidade quando fui aprovada no mestrado e por terem contribuído para minhas vitórias profissionais. 💻

# Resumo

Nos serviços da web devem ser considerados dois pontos que acabam se conflitando: i) o direito do usuário em ter sua privacidade garantida, e ii) a necessidade do mesmo em ter serviços personalizados, provendo melhor interação do usuário na Internet. As informações dos usuários, quando se tratando de privacidade, devem ser mantidas em sigilo, de forma que os mesmos não se sintam ameaçados ao utilizarem os serviços da web. Quando pensando em personalização, as informações devem ser conhecidas e filtradas, recuperando as mais relevantes para então, em conjunto com técnicas de personalização, serem utilizadas pelos sites. Este trabalho apresenta os passos seguidos para o desenvolvimento de uma ferramenta, chamada "PrivPerson", capaz de analisar os mecanismos de privacidade e personalização existentes, sendo esses: ferramentas, sites e cenários de utilização, de forma a quantificar o nível de privacidade e personalização oferecido aos usuários. Para tal mensuração, utilizou-se como base as taxonomias de classificação de privacidade e personalização, desenvolvidas neste trabalho. Estas são impostas em camadas, de modo que, a partir da observação de quais camadas o mecanismo contempla tornou-se possível quantificar o nível de privacidade e personalização oferecido. A análise de mecanismos pela ferramenta desenvolvida, objetivou fornecer aos usuários valores quantitativos que os auxiliem em seu contexto de utilização de serviços. Apresenta-se também neste trabalho, um Estudo de Caso, desenvolvido com o objetivo de identificar a utilização das taxonomias em sites, e ainda assim, comparar a avaliação manual, feita no Estudo de Caso, com a avaliação automatizada, feita através da "PrivPerson". Com base nesse estudo, observou-se que os sites, de maneira geral, não seguem uma padronização no desenvolvimento de suas Políticas de Privacidade. Baseado nisto, foi desenvolvida uma padronização para tais políticas, abordando informações que são relevantes serem apresentadas aos usuários pelos sites.

**Palavras-Chave**: Privacidade, Personalização, Internet, Web.

# Abstract

In web services two points should be considered that end up conflicting: i) the user's right to have his privacy guaranteed and ii) the need of having the services personalized, providing a better user interaction on the Internet. The user's information, with regard to privacy, should be maintained in secrecy, so that the person in question would not feel threatened when using the service. When thinking about personalization, the information should be known and filtered, in a way that makes it possible to discover the most relevant information, so that, together with personalization techniques, they can be used by the sites. This work presents the steps taken for the development of a tool, “PrivPerson”, capable to analyzing the privacy and personalization mechanisms, being: tools, sites and use sceneries in a way that quantify/measure how much privacy and personalization they offer for the users. For such measurements, a base the taxonomies of classification of privacy and personalization was used, developed in this work and imposed in layers, in a way that it is possible to quantify the privacy and personalization level of offered to the user by the observation of which layers the mechanism meditates. The mechanisms analysis for the proposed tool had the goals to supply the users quantitative values that aid the user in the context of use services. Apart from the above, a Case Study was developed with the objective of identifying the use of the taxonomies on sites, and to compare the manual evaluation on the sites, done in the Case Study, with the automated evaluation found in “PrivPerson”. With base in this study, it was observed that the sites doesn't have a standardization in the development of their Privacy Politics, based on this, a standardization was developed for such politics, approaching information that are relevant be introduced to the users by the sites.

**Keywords**: Privacy, Personalization, Internet, Web.

# Lista de Figuras

# Lista de Tabelas

# Sumário

# 1. Introdução

Com o avanço da Internet, muito se tem ganhado em relação a várias áreas de estudo. No entanto, apesar de diversos benefícios serem gerados com seu crescimento, alguns problemas também são decorrentes de sua utilização.

Ao navegar pela Internet, às vezes sem saber, os usuários deixam registros de sua navegação, que podem ser utilizados para seu benefício ou não. A coleta dessas informações pelos sites visitados pode acarretar em invasão de privacidade.

A privacidade pode ser caracterizada como o direito que o indivíduo tem de proteger sua intimidade. Cretella (1997) enfoca em sua obra o direito à intimidade visto como um direito da pessoa não ser importunada se não o desejar.

Já Silva (2002) classifica a intimidade como uma esfera secreta na vida do indivíduo, sendo o modo de ser e de viver sem interferências ou perturbações, de forma que o mesmo tenha as condições necessárias para a expansão de sua personalidade.

Apesar das necessidades de proteção à privacidade, observa-se a importância do conhecimento de algumas informações para prover serviços personalizados.

Ao longo deste trabalho é retratado o cenário onde a privacidade e a personalização encontram-se em conflito, enfatizando a necessidade do uso dessas na vida dos usuários que têm contato com a Internet. Foi feita uma abordagem sobre algumas das ferramentas de privacidade e personalização existentes, para que suas características sirvam de embasamento a alguns resultados deste trabalho.

Este trabalho tem como resultados a análise e o desenvolvimento de vários aspectos relacionados à privacidade e à personalização na web, o que inclui: i) definição de taxonomias para privacidade e personalização com a atribuição de pesos para as mesmas, objetivando-se mensurar a classificação das camadas por uma metodologia de relevância; ii) aplicação dessas taxonomias na construção de uma ferramenta para análise do nível de privacidade e personalização apresentado pelos diversos mecanismos existentes; iii) realização de um Estudo de Caso para validação de características significativas a serem apresentadas pelos sites, as quais foram embasadas nas taxonomias propostas e, iv) com os resultados observados nesse estudo, tornou-se possível o estabelecimento de um padrão para o desenvolvimento de Políticas de Privacidade.

## 1.1. Objetivo

Este trabalho teve como objetivo principal a criação de uma nova abordagem para classificação da privacidade e personalização na web, definindo taxonomias constituídas de camadas, as quais representam características relevantes de privacidade e personalização.

Outro objetivo foi o desenvolvimento de uma ferramenta, intitulada como "PrivPerson", para avaliação automatizada, com resultados quantitativos, do nível de privacidade e personalização oferecido pelos mecanismos. Para a quantificação do nível de privacidade e personalização oferecido pelas ferramentas, pelos sites e pelos cenários de utilização, foram utilizadas as camadas das taxonomias, embasadas nos pesos atribuídos a cada uma delas, tornando-se possível mensurar o nível de privacidade e personalização disponibilizada aos usuários.

De forma secundária, a ferramenta desenvolvida constitui-se como um repositório de informações sobre as ferramentas de privacidade e personalização e cenários de utilização, já que para a avaliação desses é necessário um cadastramento prévio sobre suas características, para que então, em tempo real, a análise seja feita e exibida ao usuário.

Espera-se também que soluções sejam aplicadas com vistas a tornar mais fácil ao usuário o encontro de ferramentas e sites compatíveis com seus interesses. Tornando, assim, os usuários cientes do que os cenários de utilização devem apresentar, considerando dessa forma maior confidencialidade na utilização desses mecanismos e sua satisfação.

Outro objetivo deste trabalho é oferecer mais um recurso tecnológico a pesquisadores da área e a usuários das ferramentas de privacidade e personalização, buscando assim, agregar à "PrivPerson" conceitos para atingir o estado da arte na área. Dessa forma, este trabalho visa contribuir de maneira efetiva para a área de privacidade e personalização na web, vista a demanda crescente nessa e a possibilidade de melhoria no gerenciamento das informações dos usuários que são coletadas durante sua interação com a web.

Com isso, acredita-se em um aumento no controle do usuário da web sobre sua privacidade. A ferramenta permite ainda que o usuário possa equilibrar os desejos contraditórios de ter sua privacidade protegida e, ao mesmo tempo, ter acesso a serviços personalizados.

## 1.2. Motivação

A motivação encontrada para propor este trabalho deve-se ao fato de que a privacidade pode ser colocada em risco quando o usuário necessita de serviços personalizados.

Com a ampliação do uso da Internet a privacidade está sendo ameaçada, pois tornou-se mais fácil obter os dados dos usuários. As informações obtidas podem ser utilizadas em conjunto com outras, possibilitando traçar um perfil bastante preciso do usuário, identificando suas preferências e necessidades sem que ele saiba ou veja algum perigo durante a interação com os sites.

Assim, aos poucos, a Internet vai reduzindo o direito à privacidade de uma forma a aparentar benefícios para o usuário, deturpando sutilmente a idéia de vida privada. Isso nos mostra que a privacidade é algo que está, cada vez mais, sendo diluído entre outros conceitos, como personalização, quando se tratando de Internet.

Devido ao uso indevido das informações pessoais dos usuários, aos problemas de privacidade decorrentes do uso de técnicas de personalização e à diversidade de sistemas encontrados, que tratam da privacidade do usuário e oferece personalização, fica evidente a necessidade de uma ferramenta de avaliação que possa informá-lo sobre o nível de privacidade e personalização que lhe está sendo oferecido.

Através da quantificação da privacidade e da personalização oferecida, da maior transparência quanto ao que é feito com os dados coletados e com a disponibilização de informações sobre a segurança do usuário, ele se sentirá mais seguro e confiante para interagir com os sites.

O desejo de manter a "PrivPerson" atualizada e completa motivou a busca por desenvolvê-la de forma extensível, com vistas a atender os principais pontos identificados durante o estudo do mestrado. Dessa forma, a ela podem ser adaptadas maiores funcionalidades a partir de estruturas de dados flexíveis.

Outra motivação encontrada foi a necessidade de padronização das Políticas de Privacidade, onde, baseado nisso, foi definido um padrão de Política de Privacidade a ser seguido pelos sites, de modo a atender às necessidades impostas pelos usuários.

A metodologia utilizada neste trabalho calcou-se em uma abordagem centrada no usuário, visando sua maior segurança e satisfação, a qual envolveu um conjunto de ações: i) adaptação e definição de taxonomias para privacidade e personalização; ii) experimentos com uma amostra de sites (Estudo de Caso); iii) definição de um padrão para Política de Privacidade; iv) modelagem e implementação de uma ferramenta de avaliação de privacidade e personalização; v) testes referentes à utilização da ferramenta, e vi) avaliação com usuário da ferramenta desenvolvida.

## 1.3. Organização do Trabalho

Este trabalho encontra-se organizado da seguinte forma:

O Capítulo 2 aborda os conceitos, a importância e os conflitos de privacidade e personalização, mostrando quais são as vantagens e desvantagens em se utilizar controles de privacidade quando também é necessária a utilização de personalização de serviços na Internet. Descreve também as legislações e regulamentações impostas para o uso da privacidade, enfatizando algumas regras e providências que devem ser tratadas pelos sites, para assim, aumentar a segurança do usuário.

No Capítulo 3 são apresentadas algumas técnicas de personalização para serem aplicadas a sistemas que provêem serviços personalizados aos usuários, de acordo com suas preferências e perfis.

O Capítulo 4 apresenta as estratégias que podem ser utilizadas pelos usuários para proteção de sua privacidade no mundo virtual, algumas ferramentas utilizadas para prover a privacidade e as técnicas que podem ser aplicadas para garanti-las.

A partir dos capítulos seguintes são apresentados os trabalhos desenvolvidos, bem como os resultados obtidos neste estudo, mostrando o que foi definido e desenvolvido neste trabalho.

No Capítulo 5 são apresentadas as taxonomias definidas para as camadas de proteção de privacidade e camadas de garantia de personalização (LOBATO e ZORZO, 2006).

O Capítulo 6 mostra a metodologia seguida para o desenvolvimento e a implementação de uma ferramenta, chamada “PrivPerson”, sendo feita a análise e avaliação da privacidade e personalização oferecida pelos mecanismos aos usuários.

E por fim, no Capítulo 7, são apresentados os resultados finais, decorrentes da análise e validação do desenvolvimento do trabalho.

Esses resultados englobam uma avaliação feita pelos usuários sobre a utilização e relevância da “PrivPerson” e uma comparação quantitativa. A avaliação feita pelos usuários teve como objetivo a validação da utilização da “PrivPerson”. Já a avaliação quantitativa teve como propósito a verificação da eficiência da “PrivPerson” comparando seus resultados da análise de sites com uma avaliação manual.

Tal comparação é feita através de alguns sites de comércio eletrônico. Dessa forma tornou-se possível comparar a avaliação manual, feita no Estudo de Caso, com a avaliação automatizada, feita com o uso da “PrivPerson”, para mostrar quão eficiente a “PrivPerson” se apresenta, seguindo-se da conclusão e de sugestões para trabalhos futuros.

Ainda como trabalhos desenvolvidos são disponibilizados estudos nos APÊNDICES, a fim de completar todo o trabalho.

No APÊNDICE A são disponibilizados, de forma sintética, os passos seguidos para a realização de um Estudo de Caso, que visa a avaliação por inspeção nos sites de comércio eletrônico brasileiros. Os resultados obtidos no estudo (LOBATO e ZORZO, 2007c) subsidiaram a comparação entre a análise manual com a análise automatizada, feita com o uso da "PrivPerson", apresentada no Capítulo 7.

Baseado nos resultados encontrados pelo experimento apresentado no Estudo de Caso, pode-se observar que não existe uma padronização para o desenvolvimento das Políticas de Privacidade. Dessa forma, no APÊNDICE B é apresentada uma definição de padrão para apoio ao desenvolvimento de Políticas de Privacidade, apresentando uma padronização que deve ser seguida pelos sites (LOBATO e ZORZO, 2007a) (LOBATO e ZORZO, 2007c), sendo essa constituinte das características das camadas das taxonomias.

O APÊNDICE C consiste de um complemento do Capítulo 6, onde são apresentados detalhes da implementação da "PrivPerson", bem como alguns diagramas de representatividade, os quais demonstram as ações tomadas e classes implementadas para o desenvolvimento da "PrivPerson".

# 2. Privacidade e Personalização na Web

A necessidade do usuário de ter proteção de privacidade pode afetar seu acesso a facilidades e a serviços que lhe são úteis, como os serviços personalizados na web.

De um lado, para se ter a garantia de privacidade é necessário que os dados dos usuários sejam mantidos em sigilo e, de outro, para oferecer a personalização é necessário colher esses dados para disponibilizar ao usuário melhor qualidade de serviços.

A seguir são apresentados os conceitos de privacidade e personalização e as vantagens e desvantagens de suas utilizações no contexto da tecnologia da informação, mais especificamente na Internet. São abordados também os assuntos referentes à legislação e regulamentação na Internet, de modo a trazer mais segurança aos usuários.

## 2.1. Conceitos de Privacidade

A privacidade pode ser caracterizada como o direito que o usuário tem em querer que suas informações pessoais sejam mantidas de forma segura, sem que seja possível identificá-lo enquanto navega pela web, tendo o usuário direito de poder controlar suas informações, de modo a apenas revelá-las se considerar relevante.

Com o avanço da Internet e seu crescente uso como meio de negócios e comunicações, a questão da privacidade tornou-se uma das principais preocupações em termos de segurança.

À medida que os usuários utilizam serviços na rede, deixam rastros que podem ser utilizados pelas empresas que dispõem de tecnologias suficientes para registrar as páginas visitadas pelos usuários e identificar o que foi feito em cada uma durante a visita, criando-se perfis de usuários.

Assim, da próxima vez que o usuário visitar o site, técnicas de personalização podem ser oferecidas a ele de acordo com seu perfil. Por exemplo, em sites de comércio eletrônico podem ser apresentados a ele promoções, recomendações e produtos de acordo com suas preferências.

Na web, o fato de muitas pessoas não saberem ao certo para que e o quanto de seus dados são coletados representa um grande risco à privacidade (SPIEKERMANN, GROSSKLAGS e BERENDT, 2001). Em se tratando de privacidade, o indivíduo tem o direito de permanecer sozinho, de determinar quando, como, e qual informação sobre ele será comunicada a outros. Porém, para que isso seja seguido é necessário existir leis que possibilitem exercer de fato esse direito (ISHITANI, 2003).

Assim, são criadas e descritas Políticas de Privacidade, onde é especificado o que será feito, para quem serão passadas as informações referentes aos usuários e de que forma essas serão disponibilizadas, podendo os usuários interagir ou não com esses dados mais tarde, de acordo com o que foi especificado nessas regras.

Porém, algumas vezes, essas Políticas de Privacidade podem não ser seguidas pelos sites, colocando a privacidade dos usuários em risco.

Para diminuir os riscos dos usuários terem sua privacidade invadida podem ser utilizadas técnicas e ferramentas que tratem da garantia de privacidade dos mesmos. No entanto, deve ser ressaltado que desde que o usuário tenha conhecimento do que esteja sendo feito com seus dados, e mesmo que esses sejam passados a terceiros, sua privacidade não estará sendo violada (LOBATO, BITTAR e ZORZO, 2006).

Assim sendo, é de se esperar que o respeito à privacidade seja uma das grandes preocupações no tratamento seguro da informação, sendo a discussão a esse respeito um tema delicado, visto que a autenticação e a identificação podem ser requisitos essenciais para que o acesso adequado à informação armazenada em meios eletrônicos possa ser devidamente controlado.

## 2.2. Conceitos de Personalização

Personalização é tornar algo adaptável a alguém, adequando os serviços oferecidos 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 (MAYER, 1997).

Uma das grandes vantagens da Internet é a possibilidade de receber e ter acesso a informações personalizadas, selecionadas de acordo com os interesses e as preferências de cada usuário em particular pela filtragem de conteúdos e fontes escolhidas pelos usuários.

De acordo com Koch (2003), para utilizar a personalização é necessário que algumas informações sejam obtidas de forma a utilizá-las para fins de divulgação e cálculos de estatísticas para saber a preferência dos usuários de um modo geral.

Entre as diversas áreas onde a personalização pode ser aplicada, o ambiente comercial é o que mais a utiliza, melhorando o atendimento ao usuário, permitindo que o mesmo seja diferenciado e proporcionando, assim, uma melhor interação com o site. Por exemplo, nos sites de *e-commerce*, para cada usuário que os utilizam, podem ser exibidos os produtos que melhor se enquadram a cada perfil.

Não só o usuário se beneficia com a personalização, mas também e principalmente a empresa que consegue saber qual a necessidade do mercado. Os sites que oferecem serviços personalizados conseguem, em relação a sites que não dispõem de tais serviços, uma taxa maior de conversão de visitantes em consumidores (GAERTNER e SILVA, 2006).

O acesso às informações relevantes, como preferências de usuários, torna-se imprescindível para que os dirigentes possam decidir a melhor maneira de administrar os negócios e fazer com que a organização atinja seus objetivos mercadológicos.

Pesquisas feitas demonstram que os usuários estão dispostos a fornecer seus dados pessoais e aceitam que esses sejam coletados desde que sejam utilizados para seu benefício, sob a forma de serviços personalizados. Dessa forma, os usuários podem ser auxiliados durante a navegação pelos sites, provendo uma busca mais rápida de seus interesses (GAERTNER e SILVA, 2006).

Após obter tais dados, é necessário definir quais desses são realmente relevantes, pois devem apenas ser fornecidos serviços de personalização se esses estiverem adaptados ao perfil do usuário, de modo a não trazer aborrecimentos aos mesmos expondo informações que não são de seu interesse.

Geralmente, as informações requeridas em formulários são necessárias, pois são informadas explicitamente pelos usuários suas preferências, sendo possível oferecer uma personalização mais direcionada.

Quando se observa a navegação do usuário, é necessário que sejam definidas quais informações deverão ser aproveitadas, pois o usuário pode navegar através de diversas páginas até encontrar o que realmente era procurado, e o que, será indexado ao seu perfil (TUROW, 2003).

A necessidade de personalizar os serviços oferecidos vem fazendo com que técnicas e ferramentas sejam criadas e aprimoradas com o intuito de facilitar a tarefa de personalização (ISHITANI, ALMEIDA e MEIRA JR., 2003).

Uma crítica interessante sobre os serviços de personalização é colocada por Harper (1997), onde ele levanta uma questão sobre até que ponto os serviços de notícias personalizadas não fariam com que os usuários se distanciassem da realidade, deixando de ter acesso a informações que são importantes, por essas não se adequarem ao seu perfil.

## 2.3. Conflitos entre Privacidade e Personalização

Tanto a Constituição Federal de 1988 como o Novo Código Civil brasileiro determinam que seja inviolável a intimidade, a vida privada, a honra e a imagem das pessoas. Porém, se um usuário aceita a idéia de divulgar suas informações pessoais para um site, então estaremos nos deparando com uma situação de consentimento, que não caracteriza invasão de privacidade (WARREN e BRANDEIS, 1890).

Como já citado na Seção 2.2, para haver personalização nos serviços oferecidos aos usuários, é necessário que sejam coletadas algumas informações. No entanto, depois de coletadas, não há uma maneira efetiva para controlá-las, e essa perda de controle pode acarretar em invasão de privacidade.

A garantia de privacidade e a personalização são serviços, oferecidos aos usuários, que deveriam andar unidos. Mas nem sempre isso é possível pois, se existe privacidade, existe controle dos usuários sobre seus dados e, se existe controle dos usuários, a personalização pode ser prejudicada (VOLOKH, 2000).

É necessário decidir até onde liberar informações pode ser útil, de forma a não interferir na privacidade dos usuários enquanto são disponibilizados a eles serviços personalizados.

É preciso ponderar entre o grau de importância de cada informação divulgada e o grau de confiança na segurança que o receptor pode oferecer. Porém, é difícil ter certeza se tais informações exigidas são realmente necessárias serem disponibilizadas e se o receptor é alguém em quem se possa confiar.

Baseado em observações da necessidade de privacidade e personalização, Kobsa (2002) utiliza algumas técnicas para solucionar o problema de como disponibilizar dados para os sites de forma a permitir a personalização de serviços e não invadir a privacidade do usuário, sendo elas: i) criação de Políticas de Privacidade que sejam claras aos usuários, informando o que será coletado, o objetivo da coleta e onde estarão disponíveis esses dados; ii) o próprio usuário deve decidir quais de seus dados poderão ser disponibilizados, e iii) aplicação de técnicas que garantem que o usuário não será identificado, garantindo assim sua privacidade.

Este trabalho trata a questão do conflito entre a privacidade e personalização, dando aos usuários a possibilidade de saberem o quanto desses estão sendo oferecidos a eles, quando em interação com ferramentas, sites e cenários de utilização que tratem dessas abordagens.

## 2.4. Legislações e Regulamentações para Privacidade

A Internet tem sido uma revolução para o comércio e para as transferências de dados em geral, as quais trouxeram novas oportunidades de negócios para todos, desde as grandes corporações, empresas em geral e até aos indivíduos, como usuários (GAERTNER e SILVA, 2006).

Contudo, o comércio eletrônico tem atraído crimes pela Internet, desenvolvendo-se uma nova legião de criminosos *online*, que vão desde fraudadores e *hackers* até terroristas cibernéticos. A crescente preocupação em como conduzir negócios *online* resultou no fato de que a segurança é fator fundamental para o sucesso dos negócios na Internet (KOCH, 2003).

De acordo com Elgesem (1996), o mercado está se educando com relação à segurança *online* e a maioria dos usuários agora tem a expectativa de segurança em todos os serviços utilizados na Internet, garantindo desse modo que as informações fornecidas estejam seguras durante todo o processo.

Nas seções seguintes, apresentam-se algumas regras e providências que devem ser tratadas pelos sites, de forma a aumentar a segurança oferecida aos usuários na web.

### 2.4.1. Política de Privacidade

Para disponibilizar maior segurança aos usuários, as empresas podem usar em seus sites uma Política de Privacidade. Essa política é um documento que descreve as normas seguidas pela empresa quanto à privacidade, detalhando informações sobre a utilização dos dados coletados, com quem os dados são compartilhados, como os usuários podem controlar o uso dos dados pessoais, dentre outras informações.

Várias vantagens são apresentadas com a adoção de uma Política de Privacidade adequada e clara: i) o usuário fica ciente sobre como são tratados seus dados e de que forma esses são empregados, devendo ser utilizados nos limites previstos no contrato, sob pena de descumprimento contratual; ii) inspira mais confiabilidade aos usuários que, ao tomarem conhecimento de que determinado site possui regras bem delimitadas sobre a privacidade, se sentem mais seguros quanto à divulgação ou à disponibilização de seus dados a terceiros, e iii) diminui o fornecimento de dados incorretos, devido à tentativa dos usuários de preservarem sua privacidade em ambientes suspeitos, não causando com isso problemas ao desenvolvimento do comércio eletrônico.

Ao contrário do que muitas pessoas pensam, a não adoção de uma Política de Privacidade expõe maiores riscos à empresa pois, sem a utilização dessa, o usuário poderá

julgar a coleta de seus dados como uma interceptação da transmissão de dados, o que é uma prática proibida pela nossa Constituição Federal (ROCHA, 2002).

Contudo, para a elaboração das Políticas de Privacidade deve-se considerar o tipo de empreendimento, o que será feito com os dados e quão importante eles são para a aplicação, as técnicas utilizadas na coleta de dados, elaborando assim regras moldadas ao negócio (KOCH, 2003).

As Políticas de Privacidade devem ser especificadas de forma clara e precisa, permitindo seu fácil entendimento e localização, sendo definidas cláusulas que implicam nas limitações do site e do usuário, devendo ter os sites responsabilidade de cumprir fielmente as regras impostas em suas políticas.

Qualquer modificação na Política de Privacidade deve ser comunicada previamente, por anúncios veiculados no próprio site ou por mensagens enviadas aos usuários por correio eletrônico.

Os usuários devem ter o direito de saber quais informações foram armazenadas sobre eles, bem como as fontes de coleta, caso tenham sido obtidas de terceiros, podendo corrigir ou eliminar tais informações se julgar necessário.

### 2.4.2. Leis de Proteção a Privacidade

A maioria dos usuários desejam que suas ações sejam mantidas em sigilo quando estão em contato com a Internet, porém nem sempre o sigilo é realmente colocado de forma a apenas trazer benefícios.

Uma pesquisa realizada pelo Instituto *Forrester Research*[1] mostra que em 62% das empresas americanas os funcionários acessam sites que nada têm a ver com o trabalho durante o expediente, acarretando em sérios prejuízos à empresa.

Em um estudo feito pelo *SurfWatch*[2], pode-se constatar que mais de 25% do tempo gasto pelos funcionários conectados à Internet não diz respeito ao trabalho (DOYLE, SHEVLIN e WATSON, 2004).

Esses são alguns dos exemplos em que a total privacidade pode apresentar potencial de prejuízo, tendo que ser tratado esse tema de forma a ser aplicado às necessidades impostas pelo ambiente (GARFINKEL, 2002).

[1] http://www.forrester.com
[2] http://www.surfwatch.com

Vários países discutem sobre leis que regulamentem a proteção de privacidade, de modo a aplicar uma punição a quem desrespeite essas leis. A *Eletronic Privacy Information Center* (EPIC)[3] e a *Privacy International*[4] são as responsáveis pela elaboração de um relatório informando sobre as legislações e os avanços das leis nos países, no aspecto de proteção de privacidade dos dados (GARFINKEL, 2002).

No Brasil, a situação da privacidade do usuário é preocupante devido à ausência de uma legislação clara e específica à privacidade, a qual garanta ao usuário que sua privacidade estará segura, e que se algo acontecer os infratores serão punidos (GAERTNER e SILVA, 2006).

Outro problema relacionado à privacidade no Brasil está no fato de que nem todos os sites se preocupam em utilizar Políticas de Privacidade. Quando as utilizam, alguns não trazem clareza sobre o assunto e lealdade no cumprimento das diretivas impostas nas mesmas, tendo os usuários que perder muito tempo na busca de informações (LOBATO e ZORZO, 2007c).

No entanto, apesar de não haver uma legislação que trate os crimes na web e especificamente sobre a privacidade dos usuários na mesma, existem algumas leis e projetos de leis que são uma tentativa de manter a privacidade dos cidadãos, para qualquer contexto em que ela esteja inserida (PAESANI, 2003).

Um exemplo desses é o artigo 5º da Constituição Federal do Brasil, o qual garante segurança à intimidade dos cidadãos, dizendo que:

> *Art X - " São invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurando o direito a indenização pelo dano material ou moral decorrente de sua violação".*

Outro artigo que também pode ser utilizado é o artigo 220, o qual garante liberdade de informação para os cidadãos, permitindo que eles usufruam de instrumentos informáticos, como meio de divulgação de informação e para se instruírem, sendo nesse artigo mencionado que:

*Art 220 - "A manifestação do pensamento, a criação, a expressão e a informação, sob qualquer forma, processo ou veículo não sofrerão qualquer restrição, observado o disposto nesta Constituição".*

[3] http://www.epic.org
[4] http://www.privacyinternational.org

Existem projetos, nesse contexto, que tramitam no Congresso Nacional, sendo um exemplo desses o Projeto de Lei 3.360/00. Esse projeto se refere aos crimes contra a inviolabilidade de dados e de comunicações através de computadores, os quais poderão contribuir para uma punição mais adequada aos que violam os princípios da privacidade na web, apesar de não delimitar especificamente os abusos à privacidade (AGUIAR, 2006).

As propostas da *Organization for Economic Co-operation and Development* [5] (OECD) e da *Federal Trade Commission*[6] (FTC) destacam-se no cenário internacional, objetivando regularizar a proteção de privacidade dos usuários da web, onde os princípios estabelecidos especificam de que maneira as informações pessoais dos usuários devem ser protegidas (LUCENA, 2002).

A OECD trata da proteção de privacidade dos usuários, disponibilizando e retratando documentos específicos para a segurança. Apresentam-se a seguir alguns desses princípios:

- **Princípio do Limite de Coleta:** a coleta de dados pessoais deve ser limitada e, quando essa ocorrer, deve ser feita através de meios legais;
- **Princípio da Qualidade dos Dados:** os dados pessoais devem ser autênticos, completos e relevantes para os objetivos onde serão utilizados;
- **Princípio da Especificação de Objetivo:** o objetivo da coleta deve ser especificado antes da efetivação da ação e o uso dos dados deve ser restrito aos objetivos impostos e declarados nas políticas;
- **Princípio da Limitação de Uso:** os dados coletados não podem ser divulgados ou utilizados para outros propósitos além dos especificados, exceto por uma autoridade da lei ou com o consentimento do proprietário dos dados;
- **Princípio da Segurança:** devem ser utilizadas técnicas de segurança que ofereçam uma garantia na segurança dos dados;
- **Princípio da Transparência:** deve ser criada uma política geral que trate da divulgação sobre as práticas e políticas com respeito a dados pessoais;
- **Princípio da Participação Individual:** o dono dos dados deve ter acesso a seus dados, pesquisando, visualizando e modificando-os caso julgue necessário;
- **Princípio da Responsabilidade:** um gerenciador deve ser responsável por cumprir as regras descritas, colocando em prática todos os itens acima.

[5] http://www.oecd.org
[6] http://www.ftc.gov

A FTC é uma instituição que tem por objetivo cuidar da privacidade e da vida econômica dos cidadãos americanos, auxiliando no reforço de leis a favor da segurança dos dados pessoais, vasculhando criminosos de forma a evitar fraudes em bancos. Também possibilita aos consumidores tomarem decisões de compras, permitindo assim que estejam esses melhores informados.

Sob o ato da FTC, a comissão zela contra a deslealdade e a decepção por reforçar promessas de privacidade de empresas, sobre como elas coletam, usam e asseguram informações pessoais dos consumidores (PITOFSKY, 2000).

Pela FTC são definidos alguns princípios de Práticas Justas de Privacidade, criados em 1970. Esses princípios são baseados e desenvolvidos sob uma legislação para as práticas de privacidade que protegem as informações pessoais de serem coletadas e mantidas pelo governo. Os princípios definidos pela FTC são uma síntese dos 8 princípios apresentados pela OECD, e incluem:

- **Notificação:** os sites devem manter os usuários informados sobre a coleta de seus dados;
- **Escolha:** devem ser fornecidas ao usuário opções para escolher como seus dados pessoais podem ser utilizados;
- **Acesso:** os usuários devem ter acessos a seus dados pessoais coletados, podendo atualizá-los, corrigi-los e apagá-los caso seja necessário;
- **Segurança:** os sites devem ser responsáveis e devem proteger com segurança as informações coletadas sobre os usuários.

É possível observar que ambas as propostas, OECD e FTC, se baseiam na idéia de que a privacidade está relacionada ao consentimento dos usuários sobre o que está sendo feito com seus dados. Ambas abordam assuntos sobre as formas de uso dos dados: coleta, processamento, manutenção, responsabilidade, divulgação e controle. Dessa forma, esses princípios são alguns dos assuntos que uma boa Política de Privacidade deve contemplar para oferecer mais segurança à privacidade dos usuários.

### 2.4.3. Certificados ou Selos de Privacidade

Os chamados Certificados ou Selos de Privacidade são marcas de privacidade e de confiança, desenvolvidas pela indústria de *e-commerce* e mostradas nos sites das empresas que as adquirem.

Esses informam aos visitantes que as práticas de segurança, conduzidas pelos sites, estão de acordo com o que foi proposto em suas políticas, dando ao usuário uma maior garantia de que o site cumpre as Políticas de Privacidade definidas (FRIEDMAN, KHAN JR. e HOWE, 2000).

Os Selos de Privacidade são símbolos *online* de aprovação, certificando que a declaração de privacidade do site foi examinada e testada por uma empresa que autentica os sites, como *TRUSTe*[7] ou *VeriSign*[8].

Essas empresas, chamadas de entidades certificadoras, foram criadas devido à preocupação da legislação americana em dar ao usuário uma maior garantia de que o site visitado está cumprindo as regras impostas nas leis de privacidade. A partir disso, a indústria de *e-commerce* dos Estados Unidos criou uma política de auto-regulamentação que centra o uso de selos de privacidade pelos sites (MOORES e DHILLON, 2003).

Segundo Moores (2005), para garantir que o site cumpra suas Políticas de Privacidade, as entidades certificadoras realizam uma análise no site e em suas práticas de coleta e uso das informações pessoais dos usuários, passando os sites por um processo de auditoria e avaliação, estando esses sujeitos a supervisões mesmo após a aprovação.

Se o site está cumprindo as regras definidas em sua Política de Privacidade, ele recebe o selo indicando que o mesmo passou por critérios de avaliação de privacidade, os quais garantem a sua honestidade ao efetuar transações.

A utilização de Selos de Privacidade além de trazer maior confiança aos usuários quando utilizam os sites, funciona como uma estratégia de *e-business* de sucesso, onde com o aumento da confiança dos usuários no site, há também uma possível e crescente utilização do mesmo (WANG, LEE e WANG, 1998).

No entanto, alguns sites não fazem o uso devido desses Selos de Privacidade. Em outubro de 2000 a *TRUSTe* processou dois sites, *American-Politics.com*[9] e *SurfAssured.com*[10], por uso ilegal de seu selo (MOORES e DHILLON, 2003). Outro problema, é que nem sempre os usuários reconhecem a importância da utilização dos selos pelos sites.

No próximo capítulo são apresentadas algumas técnicas que podem ser utilizadas para oferecer personalização aos usuários, de modo a oferecer serviços adequados às suas necessidades e preferências.

---

[7] http://www.truste.org/
[8] http://www.verisign.com/
[9] http://american-politics.com
[10] http://www.surfassured.com

# 3. Técnicas de Personalização

Os navegadores (*browsers*) são programas utilizados para exibir o conteúdo da requisição feita pelo usuário a um site, sendo através deles que os usuários se comunicam remotamente com os servidores onde as informações requeridas estão armazenadas.

Os navegadores 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).

Além das informações necessárias à comunicação, alguns navegadores armazenam na máquina dos usuários arquivos que servem como um identificador do usuário na rede. Esses navegadores ainda coletam informações pessoais com o intuito de prover serviços personalizados.

As informações dos usuários que são coletadas pelos sites, além de serem necessárias para disponibilizar serviços personalizados aos usuários, também são justificadas por trazerem benefícios aos sites.

De acordo com Spiekermann, Grossklags e Berendt (2001), as informações são utilizadas pelos sites para definir estatísticas, possibilitando que os mesmos conheçam as tendências do mercado.

Nas próximas sessões são apresentadas algumas técnicas que possibilitam a oferta de serviços personalizados aos usuários. Para seguir o raciocínio de personalização, onde informações dos usuários devem ser coletadas para oferecer serviços de acordo com o seu perfil, procurou-se neste capítulo não abordar, de forma enfática, as questões de invasão de privacidade.

## 3.1. Cookies

Um *cookie* é um pequeno grupo de informações, trocadas entre o navegador e o servidor de páginas, colocadas em um arquivo de texto e armazenadas pelo navegador no computador do usuário. Possui uma validade e, ao expirar é automaticamente deletadas pelo navegador.

As informações enviadas entre as páginas do servidor e o computador do usuário são realizadas através do *Hypertext Transfer Protocol* (HTTP), que tem a característica de não possuir estados, ou seja, não possuir sessões para as interações dos usuários com os sites, sendo cada interação uma transação distinta.

Assim, para solucionar a característica da ausência de estado do protocolo HTTP foram criados os *cookies*, os quais têm o objetivo de gravar dados, ações e preferências dos usuários, podendo identificá-los quando retornassem aos sites (KRISTOL, 2001).

A presença de *cookies* pode ser utilizada no contexto do comércio eletrônico para, por exemplo, saber o que está no carrinho de compras do usuário. Atualmente é considerado normal aceitar *cookies* como elementos de personalização do site, de modo a trazer conforto, inovações e, principalmente, ter um maior controle sobre os usuários.

Uma situação comum e que pode parecer estranha ao usuário, é de se encontrar armazenados em sua máquina *cookies* de sites nunca visitados, chamados de *cookies de terceiros*.

O navegador pode receber os *cookies* de terceiros quando, por exemplo, carrega uma página de um site que possui imagens de outro, que por sua vez, envia o *cookie* junto à imagem. Também podem ser repassados quando um site é assinante de algum serviço de personalização de algum outro, podendo possuir em sua página uma requisição de *cookie* desse site (KRISTOL e MONTULLI, 2000).

Os *cookies* não são necessariamente técnicas de invasão de privacidade, foram criados para melhorar a interação entre as aplicações web e seus usuários. No entanto, a forma que alguns sites encontraram para implementá-los é que pode torná-los perigosos, principalmente quando o assunto trata da privacidade do usuário.

Desde que o uso das informações coletadas por um *cookie* obedeça a algumas limitações impostas, como saber o propósito da coleta, quem está coletando, quais informações são coletadas, a quem essas serão passadas, isso não é considerado uma ameaça (SILVA, 2002).

Alguns navegadores já provêem a possibilidade de que seja mostrado um alerta antes de aceitar um *cookie* ou de barrá-lo completamente. Porém, essa prática nem sempre é usada, pois alguns *sites* necessitam de certos *cookies* para exibirem seu conteúdo de forma correta (SPIEKERMANN, GROSSKLAGS e BERENDT, 2001).

## 3.2. Web bugs

Os *web bugs* são pequenas imagens, invisíveis aos olhos humanos, inseridas em mensagens de correio eletrônico ou em sites com o objetivo de monitorar os usuários, coletando dados sobre eles.

Por serem invisíveis, os usuários acabam por carregar os *web bugs* juntamente com o restante do conteúdo de uma página. Para visualizá-los é necessário verificar *tags*[11] de imagens no código *Hypertext Markup Language* (HTML) da mensagem, que correspondem a *cookies* armazenados no computador do usuário ou da página que os contêm, ou então utilizar ferramentas para detecção de *web bugs* (MARTIN, 2003).

Apesar de, neste capítulo, terem sido deixados de lado os problemas que as técnicas de personalização acarretam, é necessário haver um equilíbrio, de modo que essas não venham invadir a privacidade dos usuários.

Ao carregar um *web bug*, que em geral pertence a um site distinto daquele com o qual o usuário está interagindo diretamente, algumas informações poderão estar sendo compartilhadas, como: endereço *Internet Protocol* (IP) da máquina que carregou o *web bug*; tipo de navegador que o carregou; URL da imagem do w*eb bug*, a qual contém a informação da comunicação entre a página web visitada e o site que coleta os dados; horário em que a imagem foi carregada.

O *web bug* coleta as informações e as transmitem ao servidor que o originou, repassando todas as informações coletadas sem que o usuário perceba o que está acontecendo, conforme mostrado na Figura 1.

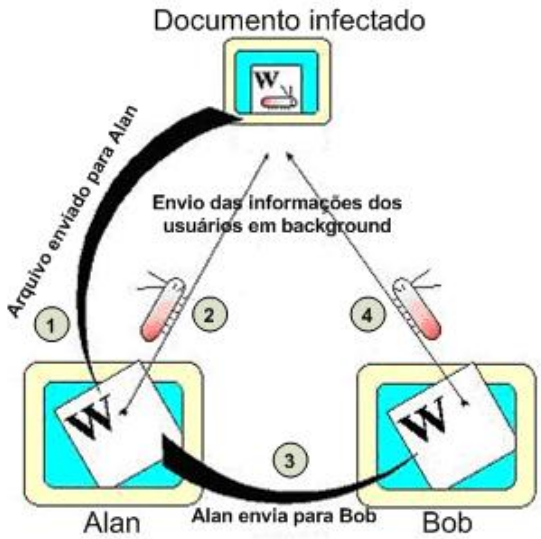


**Figura 1. Esquema feito pelos *web bugs***

Nessa figura, pode-se observar que: 1) o usuário Alan faz uma requisição a um servidor, o qual retorna a resposta da requisição. A resposta contém um documento infectado com o *web bug*. 2) Ao chegar à máquina de Alan, o *web bug* coleta informações e as envia, em *background*, ao seu servidor originário. 3) Após estar com o documento infectado com o

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

*web bug* Alan encaminha a resposta a Bob, que também torna-se infectado e o 4) *web bug* inicia seu trabalho de colher informações.

*Bugnosis*[12] e *SafeSurf* [13] são exemplos de ferramentas utilizadas para filtrar arquivos que contêm *web bugs*, de forma a oferecer segurança enquanto os usuários navegam pela Internet. A Figura 2 ilustra o processo de filtragem feito pela *SafeSurf*.


**Figura 2. Processo de eliminação dos *web bugs***

Onde: 1) o usuário envia uma requisição ao site utilizando o *proxy*[14] como intermediário, no cabeçalho dessa é adicionado o endereço IP da máquina do usuário; 2) o *proxy* retira o IP da máquina do usuário, insere na mesma o seu IP e envia a mensagem ao destino; 3) o servidor web, ao responder a requisição, indexa junto à resposta *web bugs* e a encaminha ao *proxy*, já que esse conhece apenas seu endereço IP; 4) o *proxy* filtra a resposta recebida, retirando os *web bugs* presentes e 5) a reenvia ao usuário dono da requisição.

## 3.3. Clickstream

Uma outra maneira de analisar os passos dos usuários enquanto navegam na Internet é através do uso de *clickstreams* os quais informam o caminho que o usuário percorreu durante sua navegação através de um ou mais sites.

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

Existem alguns termos utilizados para descrever a navegação na web e que são essenciais para a análise de *clickstreams* e sua melhor compreensão, sendo eles: requisição de página, sessão e visão de página (MONTGOMERY, 2003).

---

[12] http://www.bugnosis.org/
[13] http://www.safesurf.org/
[14] *Proxy* é um elemento de rede que mantém uma associação de mapeamento entre um cliente e o servidor atuando como intermediário e preservando as interações esperadas pelos clientes. Se os dados não existirem localmente no *proxy*, esse faz a comunicação e retorna os dados ao cliente (BUGLIESI e GIUNTI, 2007) .

A requisição de página é o pedido do usuário ao servidor e, para cada requisição, uma sessão é aberta ao usuário. A sessão é definida como um período de tempo durante a navegação ou uma seqüência de visões de página. Se o usuário desejar visualizar a página acessada anteriormente, será gerada uma nova visão de página e não será feita uma nova requisição, no entanto, se ele ficar certo tempo sem navegar por entre as páginas, a sessão é encerrada e uma nova requisição é feita.

Os dados de *clickstream* podem ser coletados de várias maneiras, no entanto o mais utilizado é a coleta de dados através de arquivos de *log* de servidores dos sites que estão sendo visitados (MONTGOMERY, 2003), identificando assim informações que são registradas nos *logs*, como: endereço IP, última URL, tipo do navegador, dentre outras informações relevantes à identificação do usuário.

Esses arquivos contêm todas as requisições e informações transferidas entre o computador do usuário e o servidor durante uma visita a um site. Os arquivos de *log* de servidor também podem guardar informações no identificador do *cookie* do visitante, podendo dessa forma identificar os usuários individualmente quando retornam ao site.

Ao analisar os dados de *clickstream* dos clientes as empresas podem rapidamente saber mais sobre o comportamento e os interesses dos clientes, e assim impulsionar a efetividade de seus investimentos na web.

Essa análise envolve o estudo de onde veio e para onde vai cada visitante, quanto tempo ficou no site, quantos cliques fez para ir de um lugar a outro, quanto tempo foi gasto em cada página do site, dentre outras análises específicas que podem ser feitas a partir do *clickstream*.

Um das vantagens do *clickstream* é poder ajudar o usuário indeciso em suas compras, apresentando orientações sobre os produtos de seu interesse. Essa inferência é realizada por sistemas de recomendação, que utilizam uma estrutura de preferências individuais do usuário e preferências de outros visitantes do site.

Porém, esse tipo de dados de *clickstream* é vasto em tamanho e potencialmente muito complexo, sendo necessários métodos estatísticos e de mineração de dados para tratar os dados de forma a utilizar apenas os que realmente são relevantes.

## 3.4. Data Mining

Com a utilização da Internet, grandes volumes de informações são coletados e armazenados nas bases de dados das empresas, principalmente informações referentes às

preferências dos usuários, de modo a oferecê-los serviços voltados a seus interesses. Porém, muitas vezes essas informações não são exploradas dadas as dificuldades inerentes a esse grande volume, ultrapassando assim a habilidade técnica e a capacidade humana em sua interpretação.

Essas informações passam a possuir a devida importância dentro da empresa quando são tratadas e exploradas de forma adequada, utilizando para isso as diversas tecnologias que surgem ou que são aprimoradas a cada dia.

Um exemplo dessas tecnologias é o *data mining* (mineração de dados) que tem por objetivo analisar informações em um banco de dados usando técnicas que procuram tendências ou anomalias sem o conhecimento do significado dos dados (HAN e KAMBER, 2001).

O *data mining* é uma técnica que permite extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondido nas grandes bases. Essa técnica permite que se investiguem esses dados à procura de padrões que tenham valor significativo para a empresa. Com o seu uso, é possível fazer a mineração dos dados, possibilitando que sejam colhidas nos bancos de dados, informações previamente desconhecidas ou ocultas e que sejam informações relevantes.

Diante disso, verificou-se a necessidade de explorar as grandes bases de dados, aplicando sobre esta algumas etapas do Processo de Descoberta de Conhecimento em bases de dados, chamados de *Knowledge Discovery in Databases* (KDD), a fim de identificar informações que poderiam estar "escondidas".

O processo de KDD é caracterizado como sendo composto por várias etapas interligadas. Essas etapas vão desde a definição do domínio, seleção, preparação e transformação dos dados, até a etapa de *data mining*, onde padrões podem ser "descobertos" e analisados para tornarem-se conhecimento útil (ROMERO, VENTURA e BRA, 2005).

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 quais o descobrimento será realizado. Tais dados devem ser limpos, removendo informações desnecessárias e transformados[15]. É preciso, ainda, definir a técnica e o algoritmo de *data mining* a ser utilizado.

[15] Aplicação de ferramentas de formatação a fim de gerar a base de dados de trabalho, ou seja, transformar a base de dados bruta.

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 para reconhecimento das informações relevantes podem ser realizadas novamente.

A aplicação do *data mining* torna possível comprovar o pressuposto da transformação de dados em informação e posteriormente em conhecimento. Essa possibilidade torna a técnica imprescindível para o processo de tomada de decisão.

Várias tecnologias que trabalham com manipulação de repositórios de dados podem ser utilizadas como apoio ao KDD, como *data warehouses*[16].

Fayyad, Piatetski-Shapiro e Smyth (1996) definem *data mining* como o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis.

O *data mining*, assim como as demais técnicas apresentadas neste capítulo, é importante para prover a personalização dos serviços disponíveis aos usuários, de modo que o tempo de busca pelos serviços durante a navegação na web, seja sistematicamente diminuído.

Depois de serem apresentadas algumas técnicas de personalização, no próximo capítulo são descritas as ferramentas para oferecer segurança aos usuários e proteção de sua privacidade, de modo a tornar a navegação pela web satisfatória quanto aos assuntos de qualidade de serviços e segurança.

[16] Um *Data Warehouse*, ou armazém de dados**,** é um sistema de computação utilizado para armazenar informação relativa às atividades de uma organização em bancos de dados, de forma consolidada.

# 4. Ferramentas para Proteção de Privacidade

Neste capítulo são abordadas ferramentas e técnicas utilizadas com o objetivo de manter a privacidade dos usuários protegida.

As ferramentas para garantia de privacidade possuem como característica a possibilidade de manter o sigilo sobre a identificação do usuário enquanto esse navega pela web, e também a capacidade de análise dos sites na busca por irregularidades.

Entretanto, o uso incorreto dessas ferramentas pode não permitir que a coleta de informação do usuário seja realizada, o que é fundamental para a personalização. Neste capítulo não foi dado tanto enfoque aos aspectos referentes à personalização, de modo a dar ênfase na garantia da proteção da privacidade dos usuários.

A seguir apresentam-se algumas ferramentas para proteção e garantia da privacidade, permitindo que os usuários possam navegar pela web sem serem identificados e sem que seus dados sejam coletados.

## 4.1. P3P

A privacidade é algo que a cada dia se almeja mais, principalmente porque com o passar do tempo técnicas de invasão vêm sendo desenvolvidas. Um dos métodos que podem ser utilizados para permitir a privacidade dos usuários é a Política de Privacidade, a qual descreve regras impostas e seguidas pelos sites, os seus serviços, a privacidade e a segurança oferecidas aos usuários.

O projeto da *Plataform for Privacy Preference* (P3P) é uma tentativa da *World Wide Web Consortium*[17] (W3C) de fazer uma padronização da linguagem de especificação da Política de Privacidade (COYLE, 1999).

A P3P tem como objetivo permitir com que a Política de Privacidade seja fácil de ser localizada e capaz de ser lida e entendida pelo computador de forma automática, permitindo que sites negociem com o usuário quais informações poderão ser coletadas, de que forma e onde serão utilizadas.

Para a utilização da P3P, as Políticas de Privacidade, escritas numa linguagem humana, são traduzidas para o formato XML (*Extensible Markup Language*), gerando os arquivos padrões para a P3P (CROCKER, 2005).

[17] http://www.w3.org/P3P/

Tais arquivos podem ser publicados pelo site como arquivo de referência à sua política. Isso é possível, pois a P3P possui um protocolo projetado em um formato XML, o qual permite que administradores de sites publiquem a Política de Privacidade em um formato padrão que pode ser recuperado automaticamente.

Para que esses arquivos possam ser lidos e construídos é necessário que eles obedeçam alguns parâmetros impostos sob a notação formal de uma ABNF[18]. A ABNF é uma gramática formal estendida e usada para acrescentar e melhorar a capacidade de representação proporcionada pela sintaxe XML (COYLE, 1999).

Quando o usuário interage com um site, o navegador recupera a Política de Privacidade, a qual está no formato P3P, e então avalia se essa 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, mensagens sob a forma de notificações são enviadas aos usuários, informando o ocorrido (CRANOR, BYERS e KORMANN, 2003).

Com o uso da Plataforma P3P, os usuários informam suas preferências de privacidade sem interferir nos serviços oferecidos pelos sites. Isso é possível pois apenas é avaliado se as políticas seguem as preferências dos usuários, não verificando se as diretivas impostas nas políticas são seguidas. Dessa forma, não é impedido que as funcionalidades exercidas pelos sites sejam executadas, como por exemplo, a coleta de dados durante a navegação dos usuários.

Com isso, embora a P3P forneça um mecanismo técnico de avaliação das políticas de privacidade, trazendo otimização no processo de avaliação, não traz certeza do cumprimento dessas pelos sites, sendo necessário o uso de meios específicos a essa verificação.

A especificação da P3P define: i) um esquema padrão para dados que sites podem desejar coletar; ii) um conjunto de padrões de uso, “receptores”, categoria de dados, e outras divulgações de privacidade; iii) um formato XML para expressar uma Política de Privacidade; iv) uma maneira de associar Políticas de Privacidade a sites e *cookies*, e v) um mecanismo de transporte de políticas P3P sobre o protocolo HTTP (CRANOR *et al.*, 2002).

[18] ABNF: Augmented Backus  Naur Form  for syntax specification.

## 4.2. Agentes de Privacidade

Além da P3P eliminar o trabalho que o usuário teria em analisar as Políticas de Privacidade, deixando tal análise a cargo do computador, ela possibilita às ferramentas de agente de usuário utilizá-la para informar aos usuários, através de símbolos visuais, o que está ocorrendo durante a avaliação da política.

Os chamados agentes de privacidade são ferramentas instaladas nos computadores dos usuários para proteção de privacidade. Informam usuários sobre o grau de exposição e sobre os riscos que correm em terem sua privacidade invadida.

Um exemplo de agente de privacidade é o *Privacy Bird*[19], que exibe um ícone de um pássaro no navegador do usuário, para informá-lo sobre as diretivas impostas na política do site em relação às suas preferências. Para configurar o *Privacy Bird* o usuário deve selecionar as opções de avaliação, sendo essas diferenciadas por análise em nível alto, médio ou baixo, conforme mostrado na Figura 3**Erro! Fonte de referência não encontrada.**.

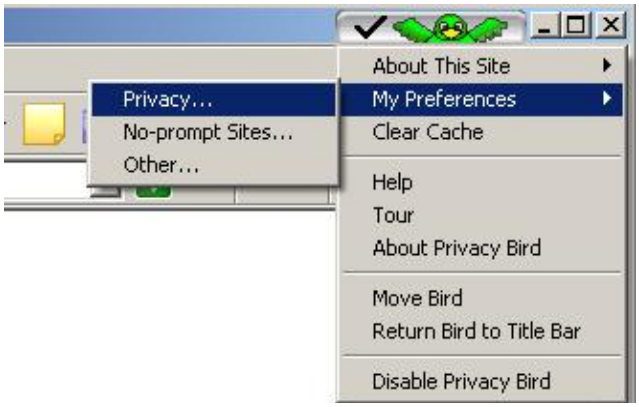


**Figura 3. Configuração do *Privacy Bird***

O problema dessa ferramenta é que a análise dos sites é feita de acordo com o que é descrito nas Políticas de Privacidade, sem conferir se realmente as regras impostas são seguidas.

Se o usuário configura o *Privacy Bird* para não permitir a coleta de dados e se na política do site é descrito que não há coleta, mesmo que essa exista, o *Privacy Bird* acredita no que foi descrito na política. Dessa forma, é informado ao usuário que pode utilizar o site pois este atende à suas exigências.

Para fazer a notificação do que está ocorrendo o pássaro muda de cor, de acordo com o grau do que foi detectado, sendo: i) vermelho, a Política de Privacidade não está de acordo com o que foi definido pelo usuário; ii) amarelo, significa que existem algumas questões que

[19] http://www.privacybird.com/

precisam ser analisadas e, iii) verde, a política atende completamente às restrições impostas pelo usuário.

Considera-se relevante utilizar tais agentes, pois esses possibilitam ao usuário ter uma idéia mais ampla do que possa estar acontecendo com seus dados, de que forma estão sendo coletados e como estão sendo tratados (ACKERMAN e CRANOR, 1999).

## 4.3. Anonimato

Com o anonimato é possível que o usuário interaja com o site sem ser identificado. Isso acontece pois o IP, assim como outras informações de identificação, são ocultadas, garantindo dessa forma, a proteção na privacidade da identidade do usuário.

O endereço IP da máquina do usuário é a principal informação que se deve proteger pois esse podem conter informações, como a localização geográfica do usuário, podendo ser utilizados para correlacionar atividades através de diferentes sites (GARFINKEL, 2002).

Alguns usuários, com medo de sua privacidade ser invadida, optam por utilizar serviços que possibilitam uma navegação de forma anônima na Internet, como o *anonymizer.com*[20] ou *the-cloak.com*[21] (SHUBINA e SMITH, 2003).

No entanto, como mostrado por Rocha (2002), nem sempre apenas essa solução é suficiente. São necessárias técnicas mais complexas, que garantam maior segurança sobre a privacidade do usuário, devido ao problema de que terceiros podem ser capazes de observar todos ou vários nós de uma comunicação.

Com o objetivo de solucionar o problema e tornar o anonimato mais eficiente, novas propostas são criadas, como por exemplo, o anonimato utilizando *proxy* (SHUBINA e SMITH, 2003). Nesse, existe um intermediário responsável por tornar o usuário anônimo, tomando a frente de suas requisições e encaminhando as repostas aos usuários sem que eles sejam identificados.

Existem dois tipos de abordagem para a utilização de um *proxy* enquanto o usuário navega anonimamente na web: *proxies* de anonimato de único nó, e *proxies* de anonimato de vários nós.

[20] http://www.anonymizer.com
[21] http://www.the-cloak.com

### 4.3.1. Proxies de Anonimato de Único Nó (Anonymizer)

Nos *proxies* de anonimato de único nó utiliza-se um *proxy* como responsável pelo envio das mensagens do remetente, sendo esse um *proxy* web que filtra todas as identificações do navegador do remetente, permitindo que ele navegue anonimamente pela web, sem revelar sua identidade ao servidor final.

Assim, o usuário envia a URL da página a ser requisitada a um servidor *proxy* e esse, por sua vez, emite uma requisição HTTP para o servidor respectivo à URL recebida. Como o *proxy* toma a frente como sendo o dono da requisição, ele mascara as requisições feitas pelos usuários, escondendo o IP dos mesmos, garantindo a segurança da privacidade do usuário, pois o único endereço IP revelado ao servidor será o endereço do *proxy* (GOLDBERG, WAGNER e BREWER, 1997).

Na Figura 4 é ilustrada a comunicação entre o usuário, *proxy* e o site destino.

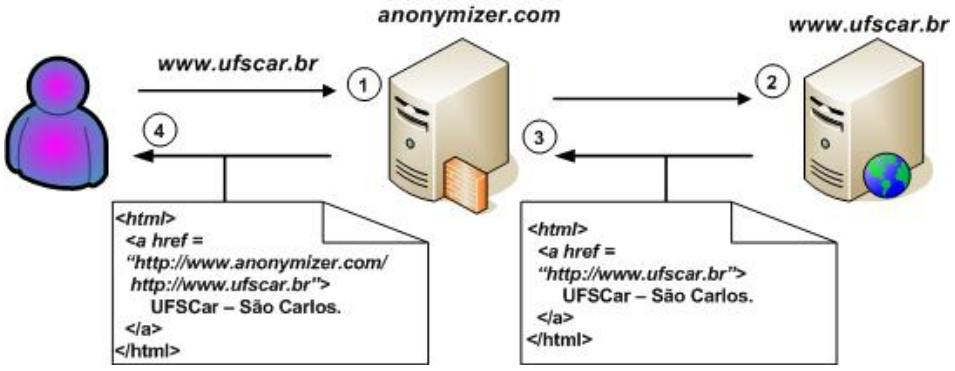


**Figura 4. Exemplo de utilização de um *proxy* de anonimato de um único nó**

Onde: 1) o usuário faz a requisição (*http://www.ufscar.br*) a um site através do *proxy* a*nonymizer.com*; 2) esse a encaminha ao site destino como sendo sua; 3) a requisição é processada e retornada do destino ao *proxy*, que supostamente é o dono da requisição e 4) o *proxy*, por sua vez, a encaminha ao verdadeiro remetente, o usuário, retornando a ele a resposta com todos os *links* reescritos apontando de volta para o *proxy* e não para o site que originalmente apontava (*http://www.anonymizer.com/http://www.ufscar.br*), garantindo seu anonimato.

No entanto, apesar de parecer suficiente para prover a navegação dos usuários pela web, sem serem identificados, existe um problema nessa abordagem. Se o *proxy* falha, não é possível continuar a navegação anônima pela Internet. Além disso, como os usuários desse serviço não são anônimos ao *proxy*, esse tipo de sistema se torna vulnerável se terceiros têm acesso ao *proxy*. Por isso, alguns *proxies* tratam desse problema encriptando o tráfego de

dados da conexão entre o usuário e o *proxy*, evitando dessa forma a ligação entre o transmissor e o receptor.

No entanto, apesar do conteúdo da mensagem estar mais seguro com a comunicação encriptada, ainda é possível saber de qual rede a requisição foi originada. Para solucionar esse problema, foram desenvolvidos os chamados *proxies* de anonimato de vários nós.

### 4.3.2. Proxies de Anonimato de Vários Nós

Os *proxies* de anonimato de vários nós utilizam uma rede chamada *mix (mix net)*, as quais são constituídas de roteadores intermediários responsáveis por transmitir a mensagem pela rede. Também têm como objetivo fazer com que a origem da requisição seja perdida, de forma a fornecer o anonimato, eliminando a ligação na comunicação entre o transmissor e o receptor.

As redes *mix* são grupos de servidores ou *proxies* responsáveis por prover o anonimato do usuário, repassando as requisições entre os vários nós da rede, podendo atrasar, reordenar, reencriptar, adicionar dados inúteis e encaminhar o tráfego passado através deles.

Dentre as várias ferramentas e tecnologias de anonimato que operam neste modelo de anonimato, ressaltam-se as seguintes: *Onion Routing* e *Crowd* (CHAUM, 1981).

#### *4.3.2.1.* Onion Routing

*Onion Routing* consiste em uma estrutura de dados em camadas, chamada *onion* (cebola), a qual é aplicada sobre a mensagem a ser enviada pela rede. Nessa são utilizados algoritmos e chaves de encriptação, usadas durante o envio dos dados, os quais são responsáveis por manter a informação segura.

*Onion Routing* se baseia em uma rede *mix,* sendo que cada nó ou roteador da rede é responsável por rotear e esconder o caminho anterior da mensagem, impedindo que o destinatário da mensagem descubra de onde tal mensagem se originou (CHAUM, 1981), como pode ser visualizado na Figura 5.

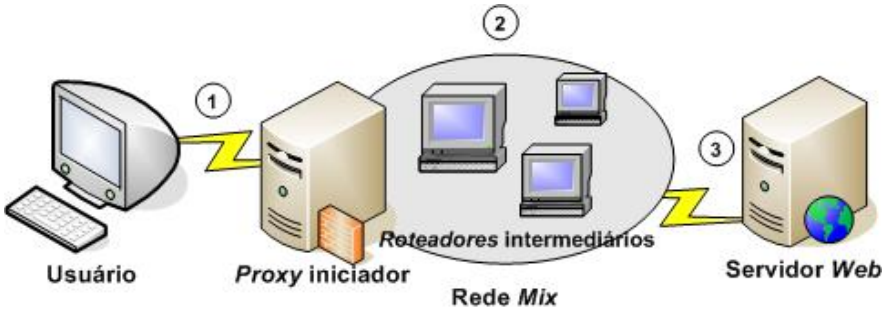


**Figura 5. Formação da Rede Mix**

Onde: 1) o usuário faz a requisição a um site, que é direcionada ao *proxy* iniciador da rede *mix*; 2) que por sua vez encaminha a requisição pela rede de nós, ou de roteadores intermediários, até 3) chegar ao destino final, servidor web.

O envio da requisição pela rede de nós acontece de forma a impedir que um espião possa associar mensagens que chegam a um nó com as que saem, já que essas são transmitidas em uma ordem diferente da ordem de chegada. Também evita que, no caso de um tráfego escasso de mensagens, sejam geradas aleatoriamente mensagens extras para outros componentes da rede.

Quando as informações vão ser transmitidas através da rede, essas são encapsuladas com diversas camadas de proteção, chamadas de *onion*, gerando a mensagem a ser enviada. Essas camadas são encapsuladas através do *proxy*, ou nó iniciador, e só depois encaminhadas à rede.

Nesse *proxy* também é definido o caminho da conexão pela rede. No entanto, há o problema da rota ser determinada no início da comunicação e um nó da rede falhar depois de a rota já ter sido definida.

Devem também ser definidas no *proxy* iniciador as informações de controle de criptografia, indicando que quando a célula contendo a mensagem chegar a cada roteador da rede, esse deve retirar uma camada do *Onion Routing*. Dessa forma, a mensagem deve chegar ao destinatário em sua forma original, contendo apenas o endereço IP do último nó do caminho, garantindo a segurança dos dados (CHAUM, 1981), como mostrado na Figura 6.

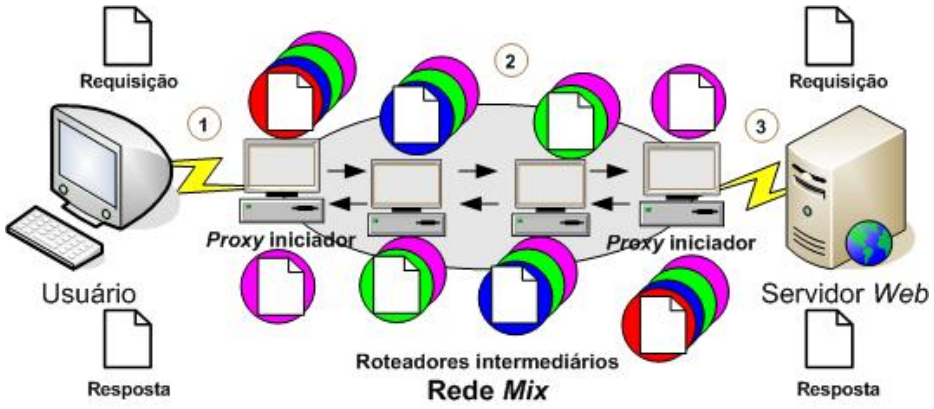


**Figura 6. Formação do caminho de comunicação em uma rede *Onion Routing***

Como o caminho a ser seguido pela mensagem na rede foi definido na hora da configuração, cada roteador da rede *onion* deve determinar sua chave pública, para decriptar completamente o *onion* que ele recebe podendo extrair informações de controle de criptografia, saber se é ele mesmo quem deveria ter recebido a célula naquele instante e a qual roteador deverá encaminhar a mensagem.

Devido ao fato das camadas serem encriptadas com chave pública, ou assimétrica, para cada roteador durante a conexão a célula aparenta ser diferente, podendo apenas a mensagem ser decriptada na ordem em que foi determinada no momento da definição da rota.

Esse plano resiste à análise de tráfego com mais efetividade que qualquer outra técnica desenvolvida para a comunicação na Internet (REITER e RUBIN, 1999).

Para os dados que se movimentam pelo caminho de volta, a construção de camadas ocorre na ordem reversa, usando diferentes algoritmos e chaves (CHAUM, 1981).

Outro ponto de vantagem nessa tecnologia é que após determinada a rota não requer um terceiro nó centralizando a conexão para o envio de mensagens.

Porém, apesar de ser tolerante a falhas, a ferramenta tem como uma de suas características as mesmas dificuldades de uma rede distribuída, e deve ser considerado um *delay* na propagação das mensagens pela rede. Também há o problema de um nó pertencente a rota falhar (SYVERSON, GOLDSCHLAG e REED, 1997).

#### *4.3.2.2.* Crowds

Essa ferramenta também tem como objetivo esconder a identidade do usuário enquanto navega pela web. Para isso, antes de realizar uma transição, o mesmo deve primeiro pertencer a uma rede *crowd* (multidão ou grupo).

Um usuário é representado em uma *crowd* por um processo *jondo*, o qual representa um participante desconhecido ou anônimo. Quando um usuário inicia o processo *jondo*, ele contacta um servidor chamado *blender* (misturador) para requisitar sua admissão na *crowd* (REITER e RUBIN, 1999).

Se o *blender* ficar inacessível não será permitida a admissão de outros *jondos*, mas existe a vantagem de que a falta do *blender* não interferirá no funcionamento da rede para envio das mensagens. O *blender* é apenas responsável por fazer as autenticações de novos *jondos* na rede, a distribuição de chaves e o relatório da sociedade atual, e não por determinar o caminho das requisições (REITER e RUBIN, 1999).

De acordo com Reiter e Rubin (1999), autores do método, após a autenticação do usuário na rede *crowd*, o *blender* seleciona um *jondo* como seu *proxy* web, especificando o nome do *host* e o número de porta em seu navegador.

Sua requisição vinda do navegador é então transmitida diretamente para esse *jondo* que, por sua vez, pode enviá-la a outros membros do mesmo grupo da rede *crowd* de forma

aleatória ou então submetê-la diretamente ao servidor final, como pode ser visualizado na Figura 7.

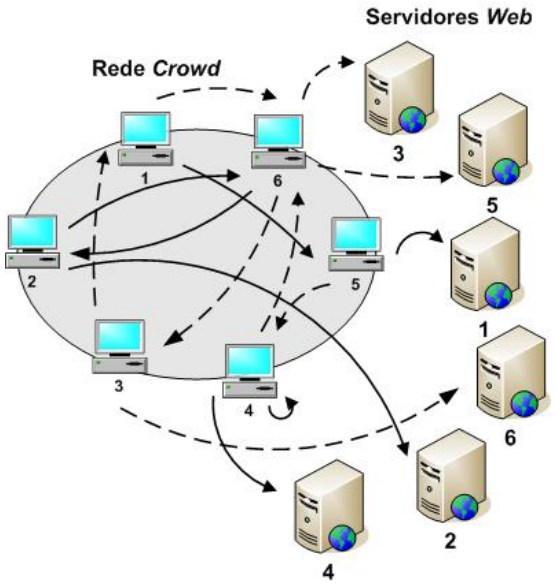


**Figura 7. Comunicação em uma rede *Crowd***

Cada membro da *crowd* possui sua própria lista sobre quais são os membros desse grupo, a cada admissão de um novo integrante, ou exclusão de um membro, essa lista é atualizada. Se um *jondo* percebe um outro que esteja falhando, ele pode eliminá-lo de sua lista, ocasionando listas desiguais, podendo esta ser atualizada quando o *blender* reportar a lista do estado atual da *crowd* para os demais, por exemplo, quando é aceito outro *jondo* na *crowd*.

Para determinar o encaminhamento ou não da requisição, o *jondo* que está de posse da requisição sorteia um número aleatório que indicará a probabilidade do envio. Se for determinado que a requisição deva ser encaminhada a algum outro *jondo*, essa será encaminhada até o destino final, senão, será submetida direto ao servidor web.

Toda a comunicação entre quaisquer dois *jondos* é encriptada usando uma chave conhecida somente pelos dois, sendo essas chaves de criptografia estabelecidas e listadas quando os *jondos* se unem à *crowd* (REITER e RUBIN, 1999).

A vantagem da rede *crowd* é que o caminho a ser seguido é definido à medida que uma mensagem é transmitida entre membros do grupo, podendo se adaptar às mudanças na rede. Ainda, o encaminhamento da mensagem aos demais é de forma aleatória, dificultando a identificação do verdadeiro remetente (REITER e RUBIN, 1999).

## 4.4. Pseudônimos

Os pseudônimos criam uma camuflagem, um apelido, para a verdadeira identidade do usuário, sendo esse um disfarce utilizado para impedir que o usuário seja identificado, permitindo o anonimato.

Ao contrário do anonimato, onde o objetivo é não se identificar, com o pseudônimo é feita uma identificação, porém não revelando a identidade real de quem o utiliza, sendo caracterizada por um anonimato de identidade real mas com revelação do pseudônimo utilizado.

Para o funcionamento do pseudônimo é necessária a utilização de um terceiro nó 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.

Apesar de não ser abordado neste capitulo a privacidade, essa é uma vantagem dessa tecnologia. Seu uso permite aos sites oferecerem serviços personalizados de acordo com perfis dos usuários, definidos através dos pseudônimos, sem identificar os usuários e sem que suas informações sejam combinadas com outros sites.

Existem diversos sistemas de privacidade baseados no uso de pseudônimos, entretanto, será apresentado o primeiro sistema, chamado de *Janus Personalized Web Anonymizer* (JPWA), pelo fato da maioria dos outros sistemas serem baseados em sua estrutura e arquitetura (MAYER, 1997).

O JPWA pode ser visto como um *proxy* que faz a comunicação entre o usuário e um site, atuando como uma entidade responsável por gerar os pseudônimos e possibilitar que o usuário os utilizem, de forma a esconder sua identidade.

Quando o usuário faz uma requisição pela primeira vez a uma página da web que necessite de uma autenticação, o sistema JPWA automaticamente reconhece a situação de login e responde mostrando no navegador do usuário um formulário de autenticação no sistema JPWA.

A partir da autenticação o *proxy* é quem se encarregará de fazer a comunicação entre o usuário e o site, mostrando ao usuário os sites que foram requisitados explicitamente por eles.

Geralmente, requisições de usuário contêm informações pessoais no cabeçalho HTTP e *cookies* enviados pelo navegador do usuário. No entanto, o JPWA filtra o fluxo de dados do navegador do usuário para o site evitando que essas informações sejam encaminhadas junto à requisição (CRANOR, BYERS e KORMANN, 2003).

A Figura 8, mostra o funcionamento do sistema JPWA, tomando como exemplo a criação e autenticação de dois usuários, Alan e Bob, em dois sites sobre esportes, *www.volei.com* e *www.tenis.com*.

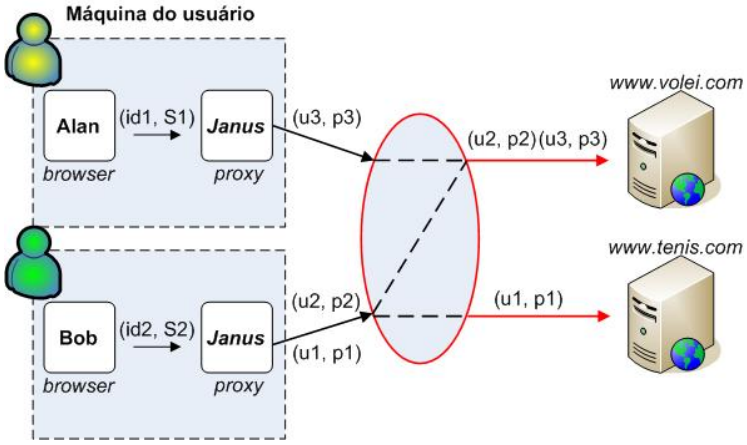


**Figura 8. Sistema JPWA, criação e autenticação automática de contas de usuário**

Conforme mostrado na figura, para fazer a autenticação do usuário Bob no sistema, esse deve informar seus dados, nome de usuário (id2) e a senha (S2), ao *proxy Janus*. Em conjunto com parte do nome do domínio do site que se quer acessar o *proxy* gera um pseudônimo de nome de usuário (u1) e senha (p1) para o site "*tenis.com*".

Se o mesmo usuário Bob deseja fazer acesso a outro site que, no entanto, tem o domínio diferente ao site acessado anteriormente, um novo pseudônimo e uma nova senha são gerados, nesse caso u2 e p2 para acessar o site *"volei.com"*.

O anonimato na web apresenta dois aspectos importantes: anonimato de conteúdo de dados e anonimato de conexão. No primeiro, a identidade do usuário não é revelada através do conteúdo do fluxo de dados entre usuários e sites, já no segundo, não é possível que o usuário seja identificado por terceiros que analisem a conexão do mesmo com o site (GOLDBERG, WAGNER e BREWER, 1997).

O sistema JPWA oferece apenas anonimato de conteúdo de dados na navegação web, deixando a desejar o anonimato de conexão. Dessa forma, é sugerido que a comunicação entre um usuário e um site se faça sobre uma rede de comunicação segura, permitindo que os grupos enviem mensagens individuais para outros e às respondam anonimamente (GOLDBERG, WAGNER e BREWER, 1997).

Para isso, o uso de *Security Sockets Layer* (SSL)[22] pode ser necessário como estratégia de proteção para as conexões, tornando o JPWA viável (GABEER *et al.*, 1998).

[22] SSL é uma camada que tem o objetivo de promover um tráfego seguro na Internet.

## 4.5. Máscaras (MASKS)

As máscaras são utilizadas com o propósito de camuflar a verdadeira personalidade dos usuários, escondendo algumas de suas características pessoais. Como apresentado por Jung (1998), o propósito da máscara é produzir uma impressão definida nos outros e, muitas vezes, embora não obrigatoriamente, isso dissimula a natureza real do indivíduo.

Na Internet as máscaras são identificações temporárias que os usuários podem assumir em suas interações, para não exporem sua verdadeira identidade (HALL e LINDZEY, 1978).

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.

Para suprir essas necessidade, apresenta-se a arquitetura MASKS (*Managing Anonymity while Sharing Knowledge to Servers*) proposta por Ishitani (2003), a qual utiliza como base o conceito de máscara e atende às necessidades impostas na utilização de máscaras pelos usuários.

O MASKS é uma arquitetura onde a idéia básica é colocar uma barreira entre os dados privados do usuário 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 (ISHITANI, ALMEIDA e MEIRA JR., 2003).

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.

Na Figura 9 é apresentada a arquitetura simplificada do MASKS com seus componentes, exemplificando a interação entre os usuários e os sites.

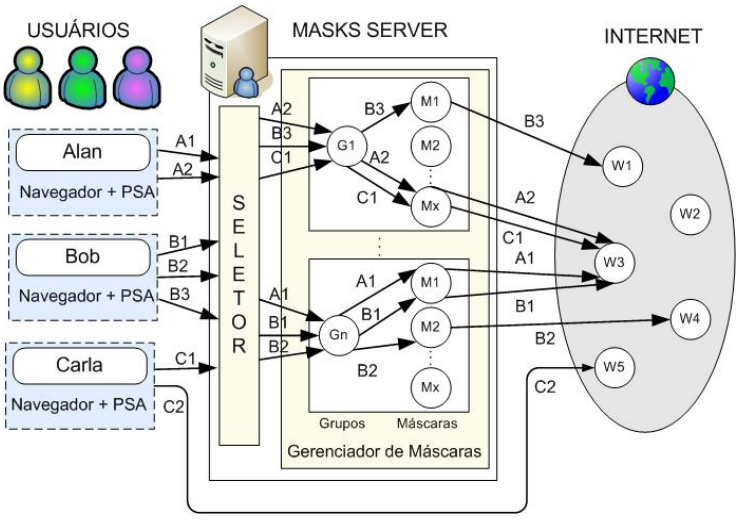


**Figura 9. Tratamento de requisições do usuário no MASKS**

O PSA é um programa intermediário entre os usuários e o MASKS SERVER que é executado junto ao navegador do usuário, 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 lhes 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 técnicas que podem causar invasão de privacidade, como os *cookies* de terceiros, *web bugs*, dentre outros (ISHITANI, 2003).

A atribuição das máscaras é baseada no conceito de grupo, sendo que cada grupo representa um tópico de interesse e, de acordo com a requisição do usuário, ele será atribuída a um grupo específico (ISHITANI, ALMEIDA e MEIRA JR., 2003).

As requisições estarão ligadas a grupos e não mais a usuários, permitindo a divulgação dos dados sobre o interesse dos mesmos e possibilitando serviços personalizados ao grupo.

O MASKS SERVER trabalha como um *proxy*, um ponto intermediário entre os usuários e os sites, sendo responsável pela criação dos grupos e pelo gerenciamento e pela atribuição das máscaras aos usuários, garantindo o anonimato.

De acordo com Ishitani, Almeida e Meira Jr (2003), o MASKS SERVER utiliza dois componentes: o seletor, responsável por definir o grupo ao qual a requisição pertencerá, e o gerenciador de máscaras, o qual é responsável por, dado um grupo, atribuir a máscara correta a cada membro desse grupo.

Assim, o servidor MASKS encaminha as requisições dos usuários usando uma máscara específica para o grupo. As respostas vindas dos sites seguem o caminho inverso e, ao chegarem nos grupos, são encaminhados aos usuários, conforme mostrado na Figura 10.

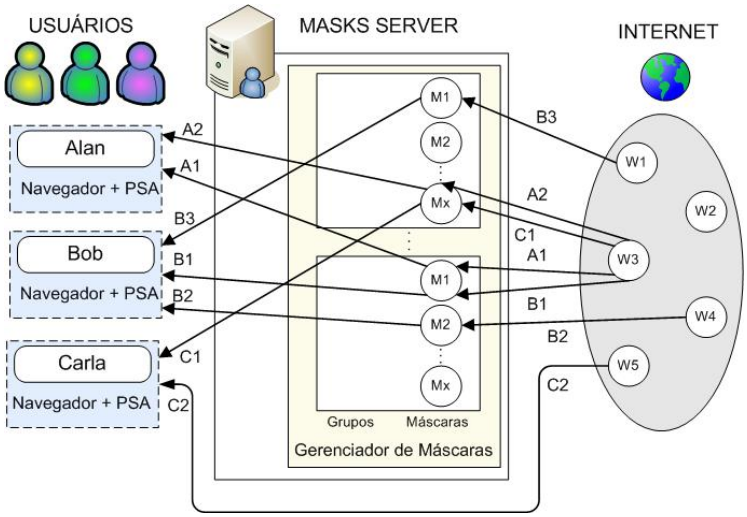


**Figura 10. Tratamento de respostas dos sites no MASKS**

O ponto principal da arquitetura MASKS é o algoritmo de designação do grupo, sendo que para cada grupo são atribuídas as requisições a páginas que estão associadas com o mesmo assunto, podendo oferecer personalização aos usuários que compõem o grupo.

## 4.6. Rede Mix

A Rede *Mix* é uma ferramenta baseada em técnicas de anonimato através de pseudônimos e tem a função de garantir a privacidade e oferecer serviços personalizados em sistemas adaptáveis ao usuário (*user-adaptative system*). Tais sistemas 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.

O sistema adaptável ao usuário é uma entidade responsável pelo gerenciamento e pela obtenção das informações do usuário, tendo por objetivo tratar as necessidades individuais de cada usuário, utilizando a Rede *Mix* para prover o anonimato. Para isso, resgatam as informações sobre as características individuais do usuário e de sua interação, transmitindo-as ao servidor que as modela (FINK e KOBSA, 2000).

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.

As informações são armazenadas em modelos individuais de usuário que são ligados ao usuário, persistindo durante diferentes sessões, como mostrado na Figura 11.

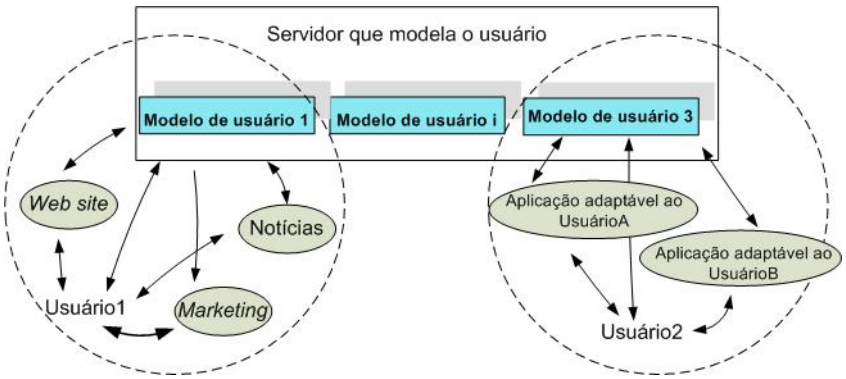


**Figura 11. Definição do modelo de usuário utilizado para criar perfis de usuário**

Após o armazenamento das informações no servidor que modela o usuário, é então aplicada a rede *crowds*, como mostrado na Figura 12.

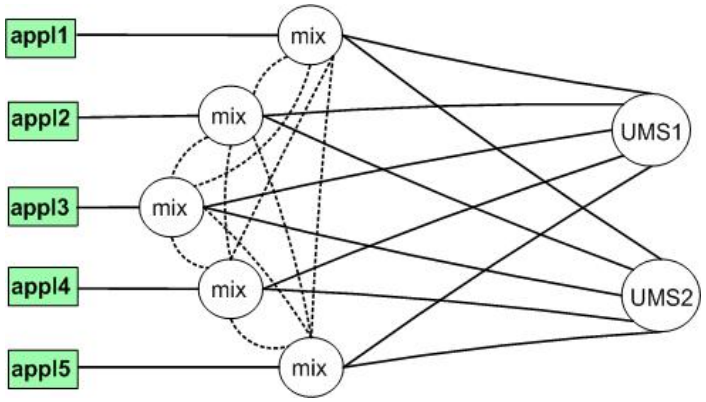


**Figura 12. Construção da Rede Mix**

O uso da rede *crowds* possibilita o envio da requisição entre diferentes nós da rede, sendo esses roteadores intermediários responsáveis por dificultar a ligação do emissor da requisição com o receptor.

Essa estrutura pode prover o anonimato dos sistemas adaptáveis ao usuário $appl_i$ e dos servidores que modelam o usuário $UMS_i$ (FINK e KOBSA, 2000).

Para prover anonimato à *appl*, escondendo a identidade do remetente, podem ainda ser utilizadas ferramentas como JPWA, *Onion Routing*, *Crowds*, dentre outros conforme mostrado na Figura 13.

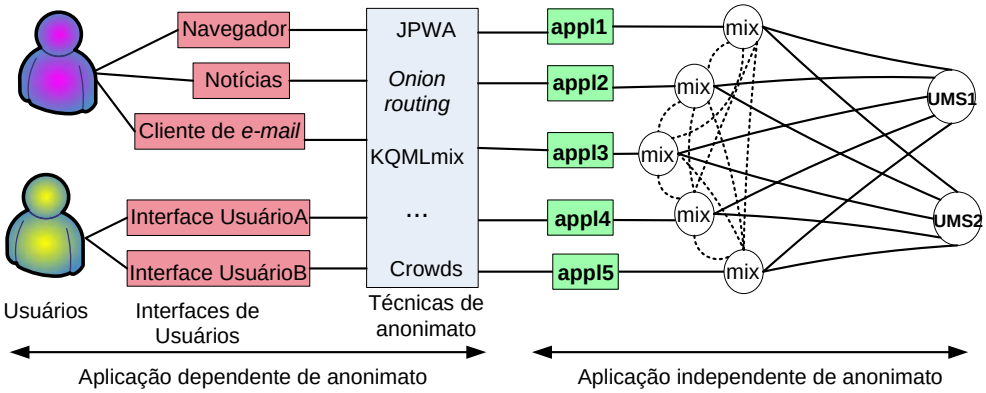


**Figura 13. Funcionamento da Rede Mix**

Além de utilizar a *crowds* como estratégia de segurança, na parte independente de anonimato, apenas o *proxy* detém conhecimento sobre o dono das informações, não sendo possível identificá-lo na rede, já que anteriormente a mensagem foi passada por outras técnicas de anonimato.

Fink e Kobsa (2000) dizem que algumas exigências devem ser seguidas para tratar da segurança em sistemas adaptáveis ao usuário, como a linguagem e o protocolo de comunicação, as quais permitem a esses sistemas trocar informações sobre o usuário com o servidor que modela o usuário. Kobsa (2001) ainda propõe o uso da linguagem KQML (*Knowledge Query and Manipulation Language*) que, de acordo com suas definições, é a

linguagem que melhor atende às necessidades impostas para segurança das informações em sua arquitetura e já tem sido utilizada em outros sistemas.

Todas as ferramentas apresentadas neste capítulo, se utilizadas corretamente podem sistematicamente garantir a segurança dos usuários e permitir a proteção de sua privacidade, de modo que esses não sejam perturbados se não desejarem.

Nos próximos capítulos serão apresentados os trabalhos desenvolvidos e os resultados obtidos, tendo como base o levantamento bibliográfico, de modo a desenvolver trabalhos realmente relevantes à privacidade e personalização oferecidas aos usuários.

# 5. Taxonomias de Privacidade e Personalização

A seguir é apresentada uma nova adaptação para a taxonomia para proteção de privacidade, e a taxonomia desenvolvida, neste trabalho, para garantia de personalização. Essas taxonomias contemplam características que devem ser apresentadas para maior segurança e qualidade dos serviços oferecidos aos usuários, e foram utilizadas para o desenvolvimento de todo o trabalho.

Essas foram utilizadas, principalmente, para tornar possível avaliar de forma quantitativa, o nível de privacidade e personalização oferecido pelos mecanismos aos usuários.

Inicialmente, pensou-se em trabalhar com uma proposta de taxonomia em níveis de privacidade e de personalização. Porém, como esclarece Millar (1995), privacidade está diretamente relacionada com a noção de consentimento, que é uma decisão completamente pessoal, não sendo possível afirmar que tudo que seja invasão de privacidade para determinado usuário será para todos os demais.

Com isso, fica claro que o tratamento em níveis não seria a melhor solução já que esses podem apresentar variações sobre as necessidades e desejos dos usuários.

De acordo com Ishitani (2003) o uso do termo “camadas” é mais correto ao contexto por não restringir que para a existência de uma determinada camada devam, obrigatoriamente, estar presentes todas as demais da estrutura.

Essas camadas irão referenciar características que devem ser apresentadas para um mecanismo prover privacidade e personalização, as quais são descritas nas próximas seções.

A partir dessa colocação e através de resultados já encontrados no levantamento bibliográfico, optou-se pela utilização de uma taxonomia em camadas para proteção de privacidade, adaptada do trabalho proposto por Ishitani (2003), e pelo desenvolvimento de uma taxonomia de personalização, também composta por camadas, a qual era desconhecida na literatura.

Quanto maior número de camadas os mecanismos (ferramentas, sites ou cenários de utilização) apresentarem, maior será a segurança oferecida e melhor será a qualidade dos serviços personalizados que são disponibilizados aos usuários, decorrendo numa maior satisfação dos mesmos, o que reflete em maiores oportunidades de negócios para as empresas, no caso da avaliação de sites.

As camadas foram dispostas de forma que uma complete a outra, ou seja, as camadas mais exteriores são aplicadas às mais inferiores, separadas por níveis de abstração, de modo a aumentar a importância da utilização das mesmas em conjunto.

## 5.1. Camadas para Proteção de Privacidade

Como nos traz Ishitani, Almeida e Meira Jr (2003) os usuários podem ter sua privacidade protegida por diferentes camadas de proteção. Cada camada é independente das demais e a existência de uma não implica que as anteriores devam existir. Contudo, quando mais de uma camada estão presentes, a organização delas seguirá sempre a mesma ordem.

Com a utilização da denominação do termo camadas, para tratar as características para a proteção de privacidade, tornou-se possível fazer uma divisão das metas que devem ser atingidas para o tratamento da privacidade. Dessa forma justifica-se a utilização da divisão em camadas, onde os mecanismos podem ser classificados em uma ou em mais camadas, dependendo da abordagem que cada uma trata.

Para tanto, foi proposta por Ishitani (2003), uma taxonomia de proteção à privacidade dividida em 6 camadas, as quais possuem responsabilidade entre o usuário e a sociedade, conforme mostrado na Figura 14.

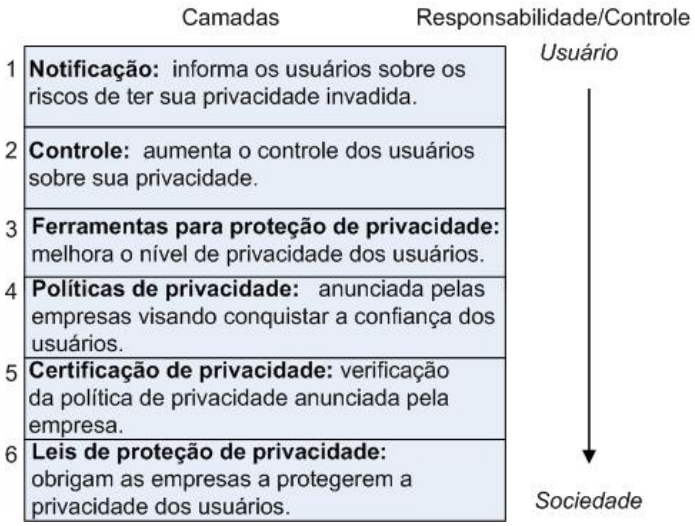


**Figura 14. Camadas de proteção à privacidade, adaptada de Ishitani, 2003**

As camadas iniciais dizem respeito às informações presentes no computador do usuário, as camadas subseqüentes encontram-se em proxies, em cada site e finalmente, as últimas camadas representam o controle de toda a sociedade regida por leis criadas pelos governos.

À medida que a análise desce através das camadas de privacidade, aumenta-se o grau de responsabilidade da sociedade e diminui a do usuário. Por exemplo, as leis de proteção de

privacidade, encontradas na camada 6 são de responsabilidade maior da sociedade do que do usuário individualmente.

Se a análise for feita na ordem inversa, a responsabilidade do usuário tende a aumentar. Por exemplo, é de responsabilidade do usuário seguir as recomendações apresentadas na camada de notificação.

De acordo com Ishitani (2003), na literatura não é mencionada nenhuma ferramenta que implemente as seis camadas de proteção de privacidade, principalmente porque é difícil implementar as duas últimas camadas, por constituírem um compromisso da sociedade e não dos desenvolvedores da ferramenta.

No entanto é possível desenvolver ferramentas que utilizam várias das camadas de proteção de privacidade. Sendo que para contemplar um maior número de camadas pelas ferramentas é importante que os pesquisadores da área estejam atentos à noção de consentimento relacionado à privacidade.

Apesar dessa taxonomia estar próxima às necessidades desse estudo, as camadas foram propostas de modo a não terem ligação entre si. Baseado nisso, decidiu-se apresentar uma nova organização dessas, de forma a ser relevante haver uma interligação entre elas.

Visa-se essa nova organização de camadas devido à necessidade de classificação de cada camada de acordo com sua importância dentro do escopo de análise de privacidade e em relação às demais camadas (LOBATO e ZORZO, 2006).

Além da determinação da ligação entre as camadas, foi definida uma nova organização, criando definições para as camadas entre o hardware e o usuário, às quais vão de funcionalidades próximas ao hardware até benefícios oferecidos ao usuário.

A Figura 15 mostra a proposta da nova arquitetura, definida neste trabalho, para as camadas de proteção de privacidade:


**Figura 15. Arquitetura das camadas de privacidade**

Verifica-se a necessidade do uso em conjunto de algumas camadas para fazer valer a existência de outras. Por exemplo, não é julgada necessária a existência da Camada 3 - Certificação de Privacidade, se não houver a Camada 2 - Política de Privacidade, a quem a certificação é aplicada.

A seguir é detalhado do que trata cada uma dessas camadas, baseado nas definições propostas por Ishitani (2003), e é mostrada uma breve explicação da necessidade de sua interligação.

- **Camada 1 - Leis de Proteção de Privacidade**: essas leis são criadas para garantir que medidas de segurança à privacidade do usuário sejam colocadas em prática e respeitadas.

Nesta camada existe a observação de que as leis se diferenciam de país para país, sendo que em alguns países elas ainda não foram definidas ou estão em tramitação. Pela inexistência dessas leis, algumas empresas podem não fornecer garantias de seus serviços de proteção à privacidade, não respeitando os direitos dos usuários.

Um problema encontrado para colocar essas leis em vigor está em se conseguir um consenso internacional para a elaboração das mesmas, já que o conceito de privacidade é extremamente dependente de questões políticas e culturais.

A Camada 1 - Leis de Proteção de Privacidade está acima das demais, pois é necessária uma lei global para o cumprimento de algumas obrigações pelos sites, porém nem sempre elas são usadas.

Se existirem as leis de regulamentação à privacidade, essa camada deverá ser aplicada às demais, permitindo assim uma maior segurança dos usuários em relação aos sites.

- **Camada 2 - Políticas de Privacidade:** as Políticas de Privacidade são uma boa estratégia para aumentar a confiança do usuário no acesso aos sites.

Utilizando esta camada será possível que os usuários conheçam a descrição das regras impostas pelos sites, informando o que é feito com os dados coletados, podendo o usuário confiar ou não em tais políticas.

Para melhorar a confidencialidade dessas Políticas de Privacidade, podem ser utilizados selos de privacidade, que fornecem uma maior garantia sobre o fiel seguimento da política pelo site.

Para maior garantia dos serviços de privacidade, pode-se utilizar a Camada 2 - Políticas de Privacidade com a Camada 3 - Certificação de Privacidade e, se possível, com a Camada 1 - Leis de Proteção de Privacidade.

Pode-se ter a camada 2 sem obrigatoriamente ter a camada 3, entretanto a garantia de segurança tende a diminuir, pois não existirá uma certificação de que as regras impostas na política estão sendo cumpridas.

- **Camada 3 - Certificação de Privacidade:** a certificação de privacidade tem como objetivo verificar, através da inspeção, se as normas descritas pelos sites em suas políticas são realmente seguidas e colocadas em práticas.

Como apresentado por Ishitani (2003) esta camada está associada à preocupação em se garantir que os sites estejam obedecendo às Políticas de Privacidade divulgadas. Para isso, a política anunciada pelo site deve periodicamente ser verificada por organizações de auditoria e grupos de privacidade.

Se os sites cumprem as diretivas impostas em suas políticas é então atribuído a eles, através de uma entidade de confiança, um selo de garantia de privacidade.

Porém, deve-se ter cuidado pois as empresas que adquirem um conjunto de dados daquelas que já possuíam esse tipo de certificação, não se sentem na obrigação de respeitar a Política de Privacidade da antiga empresa (CHELLAPPA e SIN, 2005).

Esse tipo de situação ressalta a necessidade da Camada 1 - Leis de Proteção de Privacidade para o cumprimento de tais políticas. No entanto, essa pode não ser utilizada pois em alguns países as leis não se encontram definidas.

Ainda é necessária a utilização da Camada 2 - Políticas de Privacidade em conjunto com a Camada 3 - Certificação de Privacidade, visto que, se não houver políticas não se faz necessário o uso da certificação, pois não terá a quê certificar, já que a certificação é a garantia de que o site cumpre com as diretivas impostas em sua política.

- **Camada 4 - Notificação:** a notificação é importante dentro do contexto da privacidade devido ao fato de que, muitos dos usuários não têm consciência dos riscos que podem correr em ter sua privacidade invadida durante a navegação pela web ou das vantagens em utilizar alguns serviços na Internet.

Por exemplo, alguns usuários ainda não sabem o que são os *cookies* e quais os benefícios e perdas de sua utilização, podendo bloqueá-los com medo de sua privacidade ser invadida e, com isso perderem a oportunidade de terem serviços personalizados, já que esses dependem da utilização de *cookies*. Muitos dos usuários também não sabem que cada clique em objetos do site (*links*, figuras, botões, etc) pode ser utilizado para definir um perfil detalhado sobre suas preferências.

Dessa forma, a Camada 4 - Notificação tem como objetivo alertar os usuários sobre os perigos decorrentes da navegação na web, instruí-los durante a utilização da Internet quanto às decisões que devam ser tomadas e esclarecer as dúvidas que surgirem.

Essa camada deve oferecer ao usuário acesso a diversas informações, como: o que é e para o que servem os *cookies*, as vantagens e desvantagens de sua utilização; os riscos sobre a coleta de dados; o que se ganha e o que se perde com o uso de ferramentas de anonimato; os benefícios dos serviços de personalização, dentre outras informações para notificar os usuários, principalmente, sobre acontecimentos e eventos.

Em decorrência disto, alguns navegadores oferecem a opção de definir o que se deseja proteger, já que a privacidade está relacionada diretamente com a noção de consentimento e, com isso, os usuários podem definir o que seja invasão de privacidade de diferentes formas.

- **Camada 5 - Controle:** o controle que o usuário possa ter sobre seus dados pode, consideravelmente, ser um fator de relevância à navegação na web. Essa camada inclui estratégias que permitem ao usuário ter maior controle sobre suas informações, evitando que sua privacidade seja violada.

Apesar da privacidade ser um conceito pessoal, sendo diferente para cada um dos usuários, há algumas tecnologias criadas para que a invasão da privacidade ocorra. Como exemplos podem ser citados os *cookies* de terceiros e *web bugs*, os quais têm como objetivo a coleta de dados pessoais dos usuários, com vistas à análise do comportamento desses, com ou sem os seus consentimentos.

Para esta camada, o navegador é um forte aliado pois, através dele é possível permitir ao usuário o controle da situação, de modo que o mesmo seja capaz de rejeitar ou filtrar técnicas indesejáveis de coleta de dados.

Vários navegadores já oferecem aos usuários a possibilidade de rejeição de *cookies*, entretanto, a configuração dessa opção não é trivial a usuários leigos da web. Além disso, apenas alguns dos usuários que já sofreram algum tipo de invasão ou têm conhecimento sobre os perigos da utilização da web têm interesse em utilizar esses serviços.

Devido ao pouco conhecimento do usuário sobre como controlar essa situação faz-se necessário o uso da Camada 4 - Notificação que o oriente sobre como proceder.

Também é necessário o uso da Camada 2 - Política de Privacidade, para saber sobre as práticas do site e, conseqüentemente, da Camada 3 - Certificação de Privacidade, para que uma maior garantia possa ser oferecida. Ainda faz-se necessário o uso da Camada 1 - Leis de Proteção de Privacidade para proteger o usuário, caso haja algum contravenção a seus direitos.

- **Camada 6 - Mecanismos para Proteção de Privacidade:** os mecanismos utilizados para garantir a proteção da privacidade dos usuários, desde que utilizados corretamente, são os principais responsáveis para que a segurança e privacidade sejam oferecidas aos usuários.

Nesta camada encontra-se a maior parte das ferramentas utilizadas para a proteção de privacidade, as quais possibilitam ao usuário maior tranqüilidade durante a navegação na web e a utilização dos serviços oferecidos, principalmente pelos sites.

Existem diversas ferramentas utilizadas para tratar dessa proteção, fazendo com que a comunicação entre o usuário e o site ocorra sem revelar a verdadeira identidade do usuário, e ainda assim, permitindo que o mesmo possa interagir com a web, sem prejudicar os serviços de personalização oferecidos a ele.

Como exemplos de ferramentas para proteção de privacidade podem ser citados o anonimato, mostrado na Seção 4.3, ou pseudônimos, Seção 4.4.

Tal camada pode ser utilizada em conjunto com a Camada 4 - Notificação, com a Camada 5 - Controle e, conseqüentemente, com as demais, provendo maiores funcionalidades às ferramentas de proteção de privacidade.

Deve-se ressaltar que, mesmo que existam algumas camadas que devam ser utilizadas em conjunto com outras, o maior interesse é em se desenvolver ferramentas que cubram o maior número de camadas possível, pois quando todas são utilizadas, há um aumento significativo na segurança oferecida aos usuários.

No entanto, não é apenas a contemplação de todas as camadas de proteção de privacidade que irá tornar um mecanismo dentro dos padrões para garantia da privacidade dos usuários. Também é necessário que essas sejam implementadas de maneira a não impedir que os usuários se beneficiem de alguns serviços da web, como os serviços de personalização.

Dessa forma, é necessário haver uma ponderação entre a privacidade exigida e a personalização desejada, de modo que ambas possam ser oferecidas pelas ferramentas aos usuários, seguindo as métricas relevantes às suas existências.

## 5.2. Camadas para Garantia de Personalização

A partir do estudo feito sobre camadas de proteção à privacidade, decidiu-se seguir a mesma linha de raciocínio para a definição de uma taxonomia de personalização. Essa taxonomia também é constituída de camadas, apresentando as características que são

responsáveis para prover personalização, podendo essas camadas serem utilizadas em conjunto de modo a aumentar a personalização oferecida.

A taxonomia de personalização foi criada para tornar possível a avaliação das diversas ferramentas e técnicas de personalização existentes, podendo ser classificadas em uma ou em várias camadas através de uma avaliação de forma quantificada.

Se houver a necessidade de a personalização ser feita de forma exclusiva a um usuário, ou de forma mais precisa, a utilização de mais de uma camadas da taxonomia é significantiva. No entanto, mesmo que a ferramenta não apresente todas as camadas, ou parte significativa dessas, se alguma camada estiver sendo contemplada, a personalização já estará sendo oferecida, mesmo que em um nível menor.

Também pode haver uma interligação entre as camadas das taxonomias de privacidade e personalização, de modo que se ofereça um serviço sem prejudicar o outro, havendo um equilíbrio entre a privacidade e a personalização oferecida.

Assim como na taxonomia de privacidade, as camadas que contemplam a taxonomia de personalização são dispostas de forma que estejam vinculadas a responsabilidades e controles entre o usuário e o hardware.

Deve estar implícito que para um melhor oferecimento de serviços de personalização, a ferramenta deve ser implementada de modo a disponibilizar técnicas eficientes, sendo essas de responsabilidades mais próximas ao hardware, e os usuários devem ser responsáveis pelo fornecimento de informações reais sobre suas preferências.

Assim, torna-se possível atender às necessidades dos usuários, oferecendo serviços personalizados, sem comprometer sua privacidade por meio do uso de recursos eficientes.

A Figura 16 mostra a taxonomia desenvolvida, composta por camadas de personalização, e em seguida é detalhado o funcionamento de cada uma dessas camadas, enfatizando o por que da ligação entre elas, se houver tal necessidade.

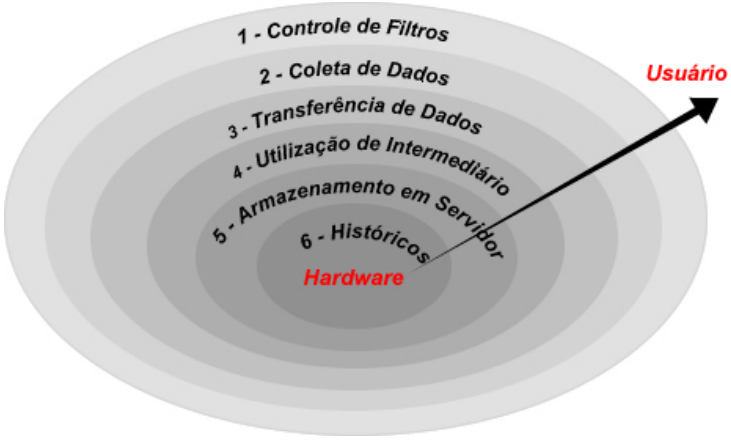


**Figura 16. Arquitetura das camadas de personalização**

- **Camada 1 - Controle de Filtros:** o controle de filtros pode ser aplicado de forma a possibilitar ao usuário configurar o que deve ser permitido em sua máquina, filtrando códigos maliciosos, como, por exemplo, *web bugs.*

Também pode ser utilizado com o objetivo de filtrar informações que chegam aos usuários, fazendo com que apenas as informações relevantes aos interesses deles sejam exibidas. Permite dessa forma o envio de conteúdo direcionado e mais atrativo, proporcionando uma maior satisfação do usuário quanto à utilização dos serviços.

Pode ser utilizada em conjunto com a Camada 4 - Utilização de Intermediário para disponibilizar tal opção de serviço aos usuários. Caso o usuário não saiba como agir para que tal camada possa ser colocada em prática, pode-se utilizar também a camada da taxonomia de privacidade, Camada 4 - Notificação, dando ao usuário informações de como proceder.

- **Camada 2 - Coleta de Dados:** a coleta de dados é um dos requisitos mais importantes para a disponibilização de serviços personalizados, pois através dela é possível ter conhecimento sobre as preferências dos usuários.

Com o uso dessa camada é possível que as ferramentas e sites coletem dados dos usuários para ser possível oferecer serviços personalizados a eles, de acordo com o perfil de cada um, adaptando os serviços disponibilizados às suas preferências.

Tal coleta de dados pode acontecer de forma implícita, onde sem que o usuário tenha conhecimento são coletados dados, ou de forma explícita com a utilização, por exemplo, de formulários, onde os usuários explicitamente informam seus dados.

Com a utilização de formulários é possível oferecer serviços personalizados mais direcionados ao perfil do usuário, já que as informações coletadas através desses são informações pessoais disponibilizadas pelo próprio usuário.

As características dessa camada são encontradas no servidor da página web, onde o usuário faz a requisição, dependendo essa camada da interação do usuário com a página para o seu funcionamento.

A Camada 2 - Coleta de Dados pode ser utilizada com a Camada 3 - Transferência de Dados para que as informações pessoais possam ser enviadas da máquina do usuário ao servidor.

- **Camada 3 - Transferência de Dados:** a transferência de dados apresenta como característica a possibilidade de troca de informações entre cliente/servidor e servidor/cliente, sendo essa uma das técnicas mais utilizadas pelos sites para oferecerem personalização aos usuários.

Um exemplo que pode ser conseguido utilizando essa camada é a possibilidade do uso de *cookies*, os quais são armazenados no computador dos usuários para permitir que dados pessoais sejam coletados.

Os *cookies* permitem que benefícios sejam oferecidos ao usuário, como a possibilidade de mostrá-lo suas últimas ações feitas durante a navegação. Permite manter conhecidos os produtos que haviam armazenados no carrinho de compras antes do usuário deixar a navegação, informar onde e o que ele estava fazendo em sua última interação, dentre outros benefícios oferecidos para facilitação de serviços e diminuição de buscas.

O uso dessa camada ainda é relevante aos sites para que esses possam identificar os usuários quando retornarem, mostrando-os assuntos e, em vezes, produtos de acordo com seu perfil. Ainda permite fazer estatísticas da navegação dos usuários, desistência e finalização de compras.

Essa camada deve ser utilizada em conjunto com a Camada 2 - Coleta de Dados para ser possível o conhecimento das informações pessoais dos usuários, e também com a Camada 4 - Utilização de Intermediário e Camada 5 - Armazenamento em Servidor, onde tais informações são armazenadas.

- **Camada 4 - Utilização de Intermediário:** as informações dos usuários podem ser armazenadas e manipuladas em máquinas intermediárias. Isso pode ser feito com a utilização de *proxies*, os quais intercedem a comunicação entre o site/usuário, usuário/site, permitindo que serviços sejam oferecidos para benefício de ambos.

Essas máquinas intermediárias podem oferecer várias vantagens aos sites e aos usuários, possibilitando a oferta de serviços personalizados, como: i) definir perfis de usuários e disponibilizá-los aos sites para que os usuários possam ser identificados; ii) permitir aos sites oferecerem aos seus usuários serviços que antes não podiam ser oferecidos, pois não estão disponíveis no servidor interno do site; iii) permitir que os usuários possam utilizá-los para que seja possível a navegação pela web sem serem identificados.

Dessa forma, são relevantes por permitirem que serviços de personalização direcionados às preferências dos usuários sejam oferecidos, aumentando a satisfação do usuário na utilização dos sites.

As informações armazenadas em máquinas intermediárias devem ser mantidas e gerenciadas apenas por pessoas que tenham permissão e que sejam de confiança da empresa. Isso se deve de modo a não infringir as camadas de privacidade, que devem ser utilizadas em conjunto para oferecer ao usuário, além de serviços de personalização, serviços de qualidade e de segurança.

Essa camada pode ser utilizada com as demais da taxonomia de personalização, de modo que a oferta de serviços de personalização seja precisa.

- **Camada 5 - Armazenamento em Servidor:** assim como na Camada 4 - Utilização de Intermediário, o armazenamento de informações no servidor dos sites é relevante, pois pode-se construir perfis de usuários e manter-se um histórico sobre as ações dos mesmos, com o objetivo de desenvolver e oferecer serviços personalizados.

Nesse caso, são previamente armazenadas as informações dos usuários nos servidores, de modo que o site possa utilizá-las para disponibilizar serviços personalizados que sejam relevantes às suas preferências.

A diferença da Camada 5 - Armazenamento em Servidor com a Camada 4 - Utilização de Intermediário, é que naquela os dados coletados são mantidos em poder dos donos dos sites e não em mãos de terceiros. Dessa forma, as informações podem ser atualizadas sempre que necessário, já que esses serviços podem ser disponibilizados aos usuários pelo próprio site, sem a intervenção de outros. Podem ainda, garantir a privacidade do usuário, já que essas informações encontram-se seguras e secretas.

As informações obtidas com a utilização dessas camadas não são de responsabilidade dos usuários e sim dos sites que as coletaram, sendo esses responsáveis por mantê-las integradas, possibilitando que os usuários possam atualizá-las sempre que necessário, conforme definido pela OECD e FTC (apresentado na Seção 2.4.2).

Assim como na Camada 4 - Utilização de Intermediário, tal camada pode ser utilizada com as demais da taxonomia de personalização, de modo a possibilitar a coleta e transferência dos dados e com isso, aumentar a precisão na disponibilização de serviços personalizados de acordo com o perfil do usuário.

- **Camada 6 - Históricos:** possibilitam o rastreamento do usuário através da observação de suas ações na Internet, sendo possível saber suas informações pessoais, as ações executadas pelo mesmo durante a interação com a web e, conseqüentemente, o caminho feito durante sua navegação.

Tais informações também podem ser utilizadas pelos gerenciadores dos sites como métricas para estatísticas e conhecimento sobre as tendências do mercado, de modo a adaptar o site ao que vem sendo buscado pelos usuários, com o objetivo de atender as vontades dos mesmos e ainda assim, visar a realização de maiores negócios.

A personalização utilizando essa camada pode ser conseguida com o uso de técnicas como c*lickstream* ou outros meios de rastreamento de informações e posterior armazenamento de históricos.

A Camada 6 - Históricos pode ser utilizada em conjunto com a Camada 3 - Transferência de Dados, possibilitando o conhecimento das ações efetuadas pelos usuários e do caminho percorrido por entre os sites durante sua navegação.

Com a definição das taxonomias de privacidade e personalização torna-se possível aproximar os mecanismos a uma garantia ideal de segurança e oferta de serviços. Assim, pode-se atender aos desejos dos usuários em ter sua privacidade protegida e ainda usufruir de serviços de personalização, de forma que a utilização das camadas não prejudique as funcionalidades dos mecanismos e de forma que as restrições impostas pelos usuários não sejam afetadas.

## 5.3. Classificação das Camadas

Nesta seção é apresentada a metodologia usada para classificação da relevância de cada uma das camadas apresentadas nas taxonomias, justificando-se a determinação de sua classificação.

É difícil avaliar o quanto os dados divulgados para os sites irão afetar as estratégias de privacidade e melhorar a personalização, ou o contrário, pois existem diversas técnicas que podem ser usadas para isso.

Entretanto, como colocado por Ishitani, Almeida e Meira Jr (2003), é possível estimar o valor da informação disponibilizada, dado um conjunto determinado de requisições, independentemente da técnica de personalização utilizada por cada site.

A classificação de relevância para as camadas foi definida de acordo com a importância delas dentro do contexto das taxonomias de privacidade e personalização e em relação às demais camadas da taxonomia em questão, seguindo-se um raciocínio linear para a atribuição dessa classificação.

Também utilizou-se para determinação da classificação da relevância das camadas o Estudo de Caso, apresentado no APÊNDICE A, através do qual verificou-se a ocorrência da utilização das características das camadas pelos sites. Com esse estudo, foi possível verificar o que é ou não realmente relevante para a segurança da privacidade e a oferta de serviços de personalização com qualidade, podendo determinar a classificação da relevância das camadas de acordo com sua real utilização.

Tal relevância seguiu o ordem de 6 a 1, de acordo com o número de camadas dispostas e tem diferença de classificação de 1 para cada uma delas, já que tem a linearidade como

característica. Assim, a camada que apresentar maior relevância, tem maior importância e, conseqüentemente, maior classificação sobre as demais.

A seguir, na Tabela 1, é apresentada a classificação da relevância das camadas da taxonomia de privacidade.

**Tabela 1. Classificação das camadas de privacidade**

| Camadas - Privacidade | Relevância dentro do Contexto |
|---|---|
| 1 - Leis de Proteção de Privacidade | 1 |
| 2 - Políticas de Privacidade | 4 |
| 3 - Certificação de Privacidade | 3 |
| 4 - Notificação | 2 |
| 5 - Controle | 5 |
| 6 - Mecanismos para Proteção de Privacidade | 6 |

Para a classificação das camadas foi verificado a importância de cada uma dentro do escopo de privacidade e em relação às demais. Isso foi feito pois existem algumas camadas que completam as outras e algumas que não são julgadas necessárias existirem se não valer a existência de outras, no sentido de trazer uma maior segurança quanto a prática que deve ser estabelecida. Por exemplo, a Camada 3 - Certificação de Privacidade, não é relevante se não existir a Camada 2 - Políticas de Privacidade, a quem a certificação é atribuída.

Na Tabela 2, é apresentada a classificação da relevância das camadas da taxonomia de personalização.

**Tabela 2. Classificação das camadas de personalização**

| Camadas - Personalização | Relevância dentro do contexto |
|---|---|
| 1 - Controle de Filtros | 1 |
| 2 - Coleta de Dados | 6 |
| 3 - Transferência de Dados | 3 |
| 4 - Utilização de Intermediário | 2 |
| 5 - Armazenamento em Servidor | 5 |
| 6 - Históricos | 4 |

Igualmente à classificação de relevância das camadas da taxonomia de privacidade, para a de personalização foi atribuído o grau de relevância de cada camada dentro do escopo de personalização, ou seja, considerando sua importância individual dentro da taxonomia e de acordo com a relação das camadas entre si.

Na próxima seção é descrito o por quê da classificação de relevância atribuída às camadas das taxonomias de privacidade e personalização. É apresentada uma quantificação, sendo essa o peso atribuído a cada uma das camadas, bem como a justificativa para a atribuição dos pesos.

## 5.4. Quantificação das Camadas

Para o embasamento de qual peso atribuir às camadas das taxonomias, foi utilizada a classificação de relevância, apresentada na seção anterior, bem como algumas observações feitas durante a realização do Estudo de Caso, APÊNDICE A. Ainda pôde-se tomar como métrica as experiências dos autores no assunto, sendo possível determinar pesos significantes a cada uma das camadas, as quais são de extrema importância à quantificação dos mecanismos.

Baseado na classificação da relevância de cada camada, essas foram quantificadas utilizando-se pesos, os quais foram delimitados podendo estar entre 1 a 10.

Dessa forma, quanto mais camadas os mecanismos apresentarem, mais próximos estarão do maior controle de privacidade e da melhor oferta de personalização, tendendo a ter peso 10, o que corresponde a 100% de garantia de privacidade ou 100% de oferta de serviços de personalização com qualidade. No entanto, não foi seguido um raciocínio de linearidade para a determinação e atribuição dos pesos às camadas.

Inicialmente todas as camadas poderiam ter pesos iguais mas, baseado na relevância, o peso de uma camada foi acrescido do número de camadas menos relevantes que ela, e nas camadas menos relevantes, os pesos foram decrementados de acordo com o número de camadas que se apresentavam mais relevantes.

Isso foi considerado, pois algumas camadas apesar de serem mais relevantes, podem apresentar um nível de relevância bem maior que outras, mesmo que a classificação de relevância tenha sido dado pela diferença de 1 entre as camadas, e então, o peso atribuído a elas deve ser maior que o peso atribuído a uma que tenha o grau de relevância diferente de 1.

Também foi levada em consideração a possibilidade de uma camada, considerada mais relevante que outra ter um grau de relevância não muito maior, devendo ser ponderada a diferença dos pesos atribuídos entre elas, já que para atribuição dos pesos não se seguiu um raciocínio linear.

A seguir, apresenta-se detalhadamente a quantificação e a classificação para as taxonomias de privacidade e personalização.

### 5.4.1. Quantificação das Camadas de Privacidade

Seguindo-se a linha de raciocínio apresentada nas seções anteriores, para a classificação e a quantificação das camadas da taxonomia de privacidade foram definidos pesos a cada uma delas.

Os pesos são responsáveis por representar a real importância das camadas dentro da taxonomia de privacidade, sendo apresentada na Tabela 3 a classificação quanto à relevância da camada e o peso atribuído a cada uma.

**Tabela 3. Quantificação das camadas de privacidade**

| Camadas - Privacidade | Relevância dentro do Contexto | Quantificação através de Pesos |
|---|---|---|
| 1 - Leis de Proteção de Privacidade | 1 | 1,0 |
| 2 - Políticas de Privacidade | 4 | 1,8 |
| 3 - Certificação de Privacidade | 3 | 1,6 |
| 4 - Notificação | 2 | 1,4 |
| 5 - Controle | 5 | 2,0 |
| 6 - Mecanismos para Proteção de Privacidade | 6 | 2,2 |

Considerada a de maior importância, a Camada 6 - Mecanismos para Proteção de Privacidade tem peso 2,2 e carrega uma grande responsabilidade, pois é nela que são definidos os mecanismos ou ferramentas disponíveis para proteção de privacidade.

A garantia está no fato de que nessa camada pode ser utilizado um servidor intermediário entre o computador do usuário e o site, de modo que esse permita que o usuário navegue anonimamente na web, sem ter informações de perfis vinculadas à sua real identidade.

Se essa camada não existisse, ficaria a cargo dos usuários ter que utilizar os sites seguindo-se medidas de segurança, podendo deixar sua privacidade ser invadida, devido à falta de experiência.

A Camada 5 – Controle, é determinada a segunda de maior importância, com peso 2,0. Através dessa, o usuário pode controlar suas informações, podendo atualizá-las, excluí-las ou até mesmo, determinar o que deve ou não ser coletado ou exibido e, caso necessário, pode barrar técnicas que julgar invasoras de privacidade, como os *web bugs* e *cookies* de terceiros.

Porém, a Camada 5 – Controle não é suficiente se não existir a Camada 4 - Notificação, pois muitos usuários não têm conhecimento sobre como bloquear e liberar essas técnicas de invasão, podendo não saber como utilizar o controle dado a eles.

A Camada 2 - Políticas de Privacidade é a próxima na classificação de relevância, com peso 1,8. Essa depende da Camada 1 - Leis de Proteção de Privacidade, para que suas práticas sejam obrigatoriamente colocadas em uso, e da Camada 3 - Certificação de Privacidade, para fornecer garantia de que as regras especificadas em suas políticas serão seguidas. Tal camada é considerada de maior relevância que as camadas 1 e 3 pois, se essa não existisse não valeria a existência das demais.

A seguir, é apresentada a Camada 3 - Certificação de Privacidade, com peso 1,6. Apesar de sua necessidade para uso em conjunto, para a Camada 2 - Políticas de Privacidade essa não é tão relevante se a camada 2 não existir pois, sem as políticas não há a quem atribuir a certificação de garantia de responsabilidade e cumprimento de regras.

A Camada 4 - Notificação tem peso 1,4, sendo responsável apenas por enviar notificações aos usuários sobre suas ações tomadas ou observações importantes. A notificação pode ser enviada sob a forma de janelas de alerta ou textos descritos nas Políticas de Privacidade e demais páginas do site, instruindo os usuários durante a navegação e informando-os sobre técnicas utilizadas pelos sites, como das necessidades de alguns *cookies* e do perigo de agentes maliciosos.

A Camada 1 é a de menor relevância, tendo peso 1,0. Isso ocorre pois nem todos os países já tem definidas Leis de Proteção de Privacidade e, dentro do escopo de avaliação dos mecanismos, essa camada não tem tanta importância, já que é mais utilizada para segurança dos usuários.

É preciso além de informar ao usuário sobre os acontecimentos e a segurança oferecida pelos sites, principalmente possibilitar que o controle da privacidade seja oferecido, na forma de elementos de privacidade, políticas e garantia de seus cumprimentos.

### 5.4.2. Quantificação das Camadas de Personalização

Utilizando a mesma metodologia para a atribuição de pesos das camadas da taxonomia de privacidade, apresenta-se a seguir a quantificação das camadas da taxonomia de personalização.

Essas foram classificadas de acordo com sua importância dentro do contexto de personalização e em relação à relevância com as demais camadas. Os pesos foram determinados sendo atribuídos os maiores valores às camadas de maior relevância.

Para a quantificação foi considerado que uma camada, apesar de ser mais relevante que a outra, pode não ser tão mais relevante, ou o contrário, e, dessa forma, uma ponderação nos pesos atribuídos foi seguida, como já mencionado na quantificação das camadas da taxonomia de privacidade.

A Tabela 4 mostra a classificação e quantificação das camadas de personalização.

**Tabela 4. Quantificação das camadas de personalização**

| Camadas - Personalização | Relevância dentro do contexto | Quantificação através de Pesos |
|---|---|---|
| 1 - Controle de Filtros | 1 | 1,0 |

| | | |
|---|---|---|
| 2 - Coleta de Dados | 6 | 2,2 |
| 3 - Transferência de Dados | 3 | 1,6 |
| 4 - Utilização de Intermediário | 2 | 1,4 |
| 5 - Armazenamento em Servidor | 5 | 2,0 |
| 6 - Históricos | 4 | 1,8 |

A Camada 2 - Coleta de Dados é a mais relevante dentro do contexto de personalização, com peso 2,2, pois com a utilização de formulários pode-se coletar informações mais específicas do usuário. Possibilita ainda a coleta explícita, o que pode evitar que a privacidade seja invadida já que o usuário tem conhecimento dessa coleta, estando a privacidade ligada diretamente ao nível de consentimento dos usuários.

Dessa forma, pode-se prover um serviço de personalização mais direcionado e que atenda às preferências dos usuários, já que tais serviços ou produtos são oferecidos baseados no perfil do usuário.

Outra vantagem encontrada com a utilização dessa camada, é que esta pode ser utilizada de modo a facilitar as demais, pois a partir de informações submetidas pelo usuário pode-se ter uma melhor análise de suas preferências, do caminho percorrido por ele, otimizando os serviços oferecidos.

Já a Camada 5 - Armazenamento em Servidor é a próxima de maior relevância, com peso 2,0. A partir do momento que as informações se encontram no servidor, ao qual os administradores do site têm permissão de acesso, há uma maior facilidade de gerenciamento das mesmas, com o objetivo de melhorar a prestação de serviços personalizados.

Para a Camada 6 - Históricos é atribuída a relevância 4, tendo essa peso 1,8, pois com a utilização de históricos é possível rastrear quase todas as ações dos usuários, definir perfis mais precisos e com isso fazer estatísticas baseadas nessas informações. Ainda com o uso dos históricos outras informações podem ser obtidas, como por exemplo verificar o que leva os usuários a tomar certas decisões durante a navegação pela web, como o abandono de carrinhos de compra.

O conhecimento das ações dos usuários e de informações referentes à sua navegação é vantajoso aos sites, pois esses podem desenvolver técnicas e serviços para atraí-los e melhorar a interação, evitando que a navegação seja abandonada e as ações não sejam efetivadas.

Essas melhorias podem ser feitas através de técnicas de personalização mais direcionadas, já que se tem um perfil bem definido dos usuários através dos históricos de sua navegação.

Com peso 1,6 aparece a Camada 3 - Transferência de Dados, definida com relevância 3 devido à facilidade de sua existência e eficiência, pois a partir do momento que se inicia

uma navegação pela Internet, informações são trocadas entre cliente/servidor e servidor/cliente.

Um bom exemplo da transferência de dados é o uso de *cookies*. Esses podem ser gerados e armazenados na máquina do usuário para coleta de informações pessoais, sendo essas informações utilizadas para a definição dos perfis de usuário e para oferecer serviços personalizados.

É interessante que ao usar essa camada, os mecanismos também contemplem algumas camadas da taxonomia de privacidade, de modo a garantir que a privacidade do usuário não seja invadida.

A Camada 4 - Utilização de Intermediário tem peso 1,4, onde o funcionamento dos serviços de personalização está relacionado com a possibilidade de utilização de um servidor intermediário ou *proxy*, o qual intercepta a comunicação entre o usuário e o site.

O *proxy* é o responsável por agilizar algumas ações dos usuários, por se passar por ele na comunicação com o site, e por retornar as respostas personalizadas aos usuários.

A qualidade dos serviços oferecidos com o uso dessa camada está ligada à maneira como o usuário é representado pelo *proxy*, pois se, por exemplo, os usuários forem atribuídos a grupos, um pouco da precisão da personalização oferecida pode ser perdida, já que não se tem perfis únicos para os usuários e sim perfis de grupos nos quais os usuários encontram-se inseridos.

E, finalmente, a Camada 1 - Controle de Filtros é a de menor relevância, tendo peso 1.0, pois se o usuário utilizar esse serviço de personalização e bloquear muitos dos serviços que podem ser oferecidos a ele, a personalização poderá ser prejudicada.

Essa camada é necessária, pois se utilizada pelos mecanismos de forma correta, pode evitar que informações que não se enquadram ao perfil de um usuário específico sejam enviadas. Isso evita causar aborrecimentos aos usuários e provê facilidades de personalização. No entanto, como às vezes os usuários não têm muito conhecimento sobre o que cada mecanismo representa, por medo de sua privacidade ser invadida, acabam por bloquear os serviços.

## 5.5. Conclusões

Através da definição das taxonomias de privacidade e personalização e da classificação e quantificação de suas camadas, baseadas na relevância que elas apresentam,

pode-se observar o quão importante essas são para a parametrização dos valores de privacidade e personalização que são oferecidos pelos mecanismos aos usuários.

Utilizando a quantificação das camadas é possível fornecer aos usuários dados concretos sobre o nível de privacidade e personalização que é oferecido e ainda apresentar quais dessas camadas os mecanismos contemplam, sendo essas características importantes à privacidade e a personalização.

Como medidas de comparação entre a relevância das camadas e exemplificação do uso dessas para mensuração do nível de privacidade e personalização, foram avaliados três mecanismos, podendo esses ser: ferramentas de privacidade e personalização, sites ou cenários de utilização.

Tais mecanismos, denominados Mecanismo X, Y e Z, foram tomados como mecanismos exemplos para este estudo. Esses apresentam diferentes camadas das taxonomias, e com isso, diferentes níveis de privacidade e personalização são oferecidos aos usuários.

Os gráficos apresentados a seguir, foram disponibilizados apenas como métricas de exemplificação. Tal exemplificação foi feita de modo a ressaltar a importância da classificação das camadas e dos pesos atribuídos a elas. De acordo com a relevância das camadas dentro das taxonomias e, com o grau de relevância de uma camada para outra, podendo uma camada ser muito mais relevante que as outras, ou com um nível de relevância aproximado.

No primeiro exemplo, apresentado na Figura 17, para avaliação da privacidade são retratados dois mecanismos, chamados de Mecanismo X e Mecanismo Y, os quais apresentam as camadas 1, 2, 3 e 4, e, 1, 2, 3 e 6, respectivamente.

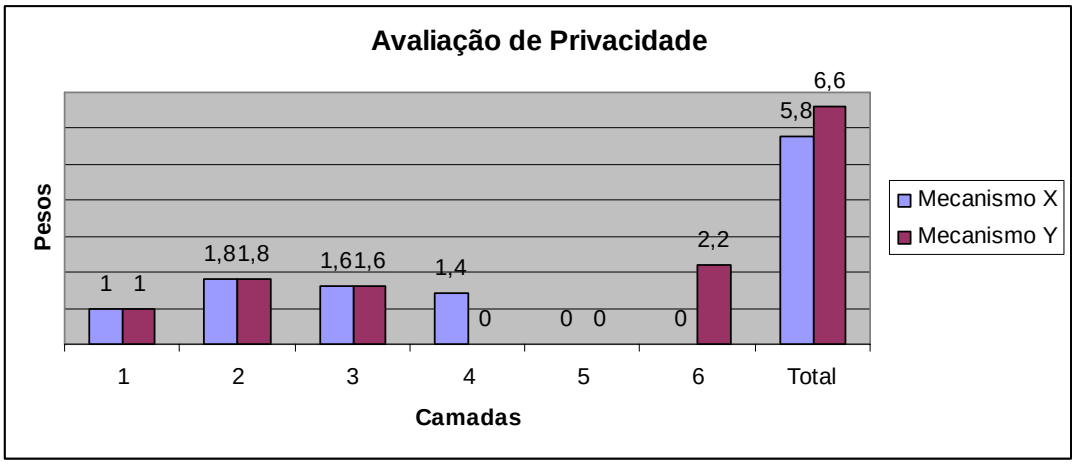


**Figura 17. Avaliação exemplo do nível de privacidade oferecida**

Pode-se observar neste gráfico que apesar dos mecanismos abordarem o mesmo número de camadas, o nível de privacidade oferecido por eles é diferente, tendo apresentado o

Mecanismo X peso total 5,8, e o Mecanismo Y peso 6,6, ou seja, é disponibilizado aos usuários um nível de 58% e 66% de privacidade possível de ser oferecida aos usuários.

Para os mesmos mecanismos do exemplo, foi avaliado o nível de personalização oferecido, onde os respectivos mecanismos apresentam as camadas 2, 3, 4, 5 e 6, e, 2, 3, 5 e 6 da taxonomia de personalização. Na Figura 18 o resultado de tal exemplificação é mostrado.

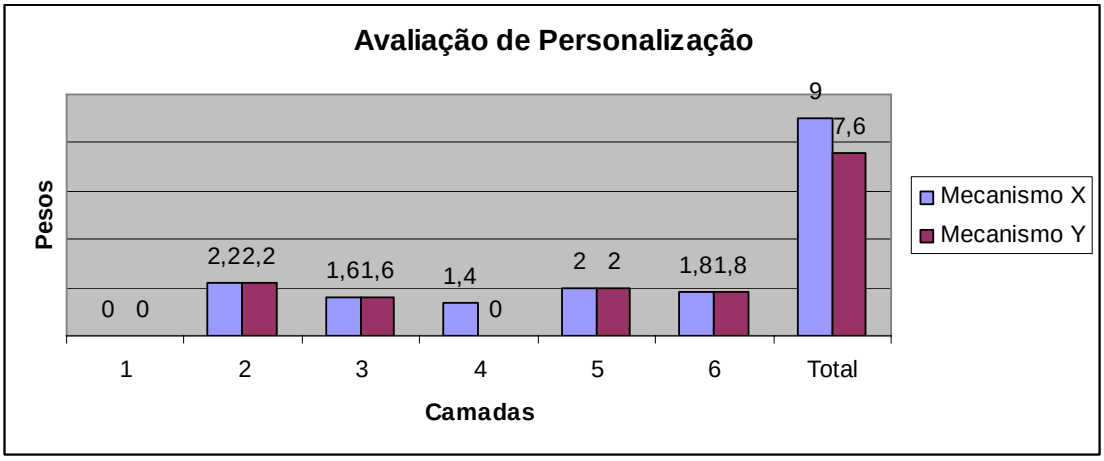


**Figura 18. Avaliação exemplo do nível de personalização oferecida**

Nesta figura ambos os Mecanismos não apresentam a Camada 1 - Controle de Filtros e o Mecanismo Y não apresenta a Camada 4 - Utilização de Intermediário, o que faz com que o Mecanismo X apresente 90% de nível de personalização, enquanto o Mecanismo Y apresenta 76%.

No próximo exemplo é mostrada a avaliação feita para o Mecanismo Z, onde é feita a avaliação do nível de privacidade e personalização que o mesmo apresenta. Tal mecanismo contempla as camadas 1, 2, 3, 4, 5 e 6 da taxonomia de privacidade e 1, 2, 4 e 5 da taxonomia de personalização, como pode ser visualizado na Figura 19.

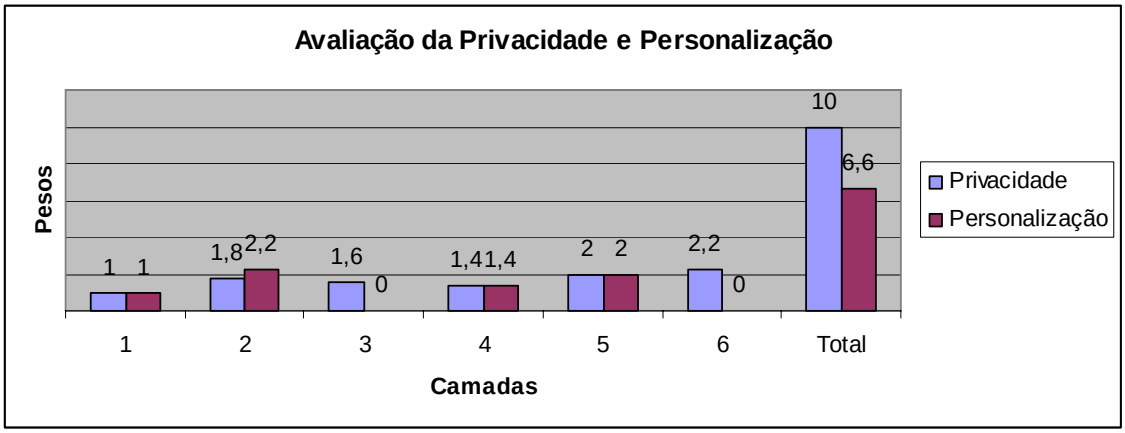


**Figura 19. Avaliação exemplo da privacidade e personalização**

Observa-se na Figura 19 que a privacidade é oferecida em um nível bem maior que a personalização. O mecanismo apresenta todas as camadas da taxonomia de privacidade,

oferecendo 100% de garantia de privacidade, baseado nas métricas de avaliação da ferramenta, e 66% de nível de personalização.

A avaliação dos mecanismos foi baseada nessas taxonomias, pois privacidade e personalização são assuntos subjetivos e em um alto nível de abstração, dependentes de regiões e crenças, e o que pode ser invasão de privacidade para uns pode não ser para outros. Dessa forma, a avaliação não foi feita em uma privacidade ou personalização geral, e sim, de acordo com as camadas que as taxonomias contemplam.

Para validar a importância das camadas das taxonomias de privacidade e de personalização, essas foram aplicadas em um Estudo de Caso, APÊNDICE A, para que pudesse verificar se as camadas eram utilizadas mesmo antes de serem conhecidas. Neste estudo foram avaliados manualmente 33 sites de comércio eletrônico, com vistas a verificar as características contempladas.

Como as camadas representam conceitos amplos e subjetivos dentro da análise de sites, essas foram divididas em itens, os quais representam as mesmas características das camadas, mas que, no entanto são transcritos de uma maneira mais usual e simplificada. Isso foi feito de modo a caracterizar um refinamento computacionalmente possível de ser analisado, sendo nomeados como as características que os sites podem apresentar. Assim, os itens são constituídos de algoritmos que juntos representam toda a subjetividade das camadas.

O Estudo de Caso é de grande importância para a validação dos resultados de todo o trabalho desenvolvido, tendo sido utilizado para: i) comprovação da utilização das camadas das taxonomias pelos sites; ii) determinação da nova organização da taxonomia de privacidade; iii) possibilidade de quantificação das camadas, observadas a utilização dessas nos sites; iv) desenvolvimento de um padrão para a definição das Políticas de Privacidade apresentadas pelos sites, pois foi observado que nenhum padrão é seguido, o que dificulta o entendimento dos usuários; v) comparação da avaliação manual com uma avaliação automatizada, feita pela ferramenta “PrivPerson”, de forma a validá-la e verificar sua eficiência. O Estudo de Caso pode ser verificado no APÊNDICE A.

# 6. “PrivPerson” – Ferramenta de Avaliação

Para solucionar as questões conflitantes entre privacidade e personalização, diversas estratégias, técnicas e ferramentas podem ser empregadas, de forma a tornar as informações dos usuários seguras e ainda oferecer personalização com o uso de serviços disponíveis na web.

Neste trabalho foi desenvolvida uma ferramenta, chamada “PrivPerson”, a qual avalia de forma mensurável, o nível de privacidade e de personalização oferecidos na utilização de alguns mecanismos como: i) ferramentas de privacidade e personalização; ii) sites de interesse do usuário, e iii) cenários de utilização, sendo esses representados por ações feitas pelos usuários em interação com a web.

Para a avaliação dos mecanismos pela “PrivPerson” são utilizadas as taxonomias de privacidade e personalização, sendo verificados quais camadas o mecanismo em análise apresenta e então, de acordo com o peso de cada camada, é retornado o nível de privacidade e personalização encontrado.

A seguir é apresentada a “PrivPerson” e a metodologia seguida para seu desenvolvimento, exemplificando o que foi utilizado para a criação da mesma e como essa avalia o grau de privacidade e personalização que os mecanismos apresentam, seguindo de suas interfaces. Detalhes de implementação, bem como diagramas que representam as ações tomadas para o desenvolvimento da “PrivPerson” e as classes desenvolvidas são mostrados no APÊNDICE C.

## 6.1. Descrição da Ferramenta

Quando técnicas e ferramentas de privacidade e personalização são utilizadas em conjunto, torna-se possível disponibilizar aos usuários serviços para garantia de privacidade e oferta de personalização. No entanto, é difícil haver um controle por parte dos usuários sobre o que esses mecanismos dizem oferecer e o que realmente é oferecido.

Dessa forma, a “PrivPerson” disponibiliza ao usuário o nível de privacidade e personalização que é oferecido pelos mecanismos, tornando-o consciente sobre os riscos e os benefícios apresentados durante o uso desses.

Uma representação do modo de funcionamento da “PrivPerson” é apresentada na Figura 20, onde deve ser informado pelo usuário o que deverá ser avaliado e então a análise é feita pela “PrivPerson” e a resposta é retornada de forma clara e precisa.

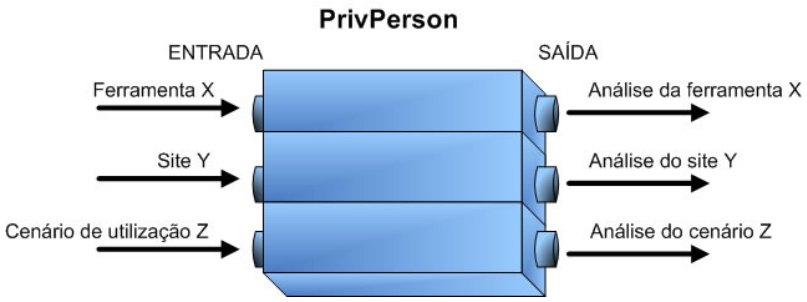


**Figura 20. Estrutura da "PrivPerson"**

Observa-se, em relação ao processo de identificação de privacidade e personalização, referente ao que as ferramentas apresentam, ao que os sites possibilitam e ao que os usuários desejam, foram utilizadas as seguintes metodologias e posterior avaliação automatizada para quantificar o nível de privacidade e personalização apresentado. A "PrivPerson" precisa inicialmente de uma entrada com a opção desejada do mecanismo a ser avaliado, para análises do tipo:

- Uma ferramenta que trata de privacidade e personalização;
- Um endereço válido de um site;
- Um cenário de utilização, sendo esse as ações que os usuários podem fazer durante a interação com a web.

Para a avaliação das ferramentas e dos cenários de utilização é necessário que o usuário selecione qual ferramenta ou cenário deseja avaliar, dentre as opções já cadastradas na "PrivPerson". No entanto, tais mecanismos podem ser atualizados, inseridos e apagados, tornando dessa forma a "PrivPerson" adaptável a outras ferramentas e cenários.

A "PrivPerson" utilizou as taxonomias de privacidade e personalização e a quantificação das camadas, apresentadas no Capítulo 5, para mensurar a resposta da análise dos mecanismos retornada aos usuários.

Conhecidas as características dos mecanismos, foram então cadastradas na "PrivPerson" informações de quais camadas são contempladas pelos mesmos e, já tendo o peso para cada camada das taxonomias, tornou-se possível fazer a análise de maneira correta e retornar ao usuário o nível de privacidade e personalização oferecido.

Tais informações são referentes a quais camadas as ferramentas e cenários contemplam e às características a serem verificadas na avaliação dos sites. Isso é feito de modo que, com o conhecimento dessas informações, tornou-se possível a análise e a posterior quantificação dos resultados aos usuários.

Para verificar a existência do uso das características, apresentadas nas taxonomias, pelos mecanismos avaliados, foi utilizada a definição de itens para tal inspeção. Tais itens, são

apresentados no APÊNDICE A. Foi necessária a verificação através desses itens pois esses são mais descritivos e não tão subjetivos quanto as camadas, estando mais próximos ao entendimento humano e à verificação computacional.

O cadastro prévio de quais camadas os mecanismos apresentam é feito apenas para as ferramentas e os cenários de utilização, pois essas apresentam informações e características que não mudam constantemente.

Para a avaliação dos sites não houve cadastro de informações referentes aos mesmos devido ao grande número de sites existentes, o que torna o cadastramento inviável, e devido à possibilidade de constantes modificações, o que poderia em vezes tornar falha a avaliação. Dessa forma, a verificação de quais camadas os sites contemplam é feita em tempo real, podendo a resposta da análise ser diferente para as análises de um mesmo site em faixas de tempo distintas.

A principal funcionalidade atribuída à "PrivPerson" é a possibilidade de avaliação dos mecanismos, fazendo a análise do nível de privacidade e de personalização oferecido. No entanto, essa apresenta outras funcionalidades e configurações, as quais foram desenvolvidas para facilitar sua utilização e serão descritas ao longo deste capítulo.

Na Figura 21 é mostrada a tela inicial da "PrivPerson", onde algumas informações são disponibilizadas aos usuários.

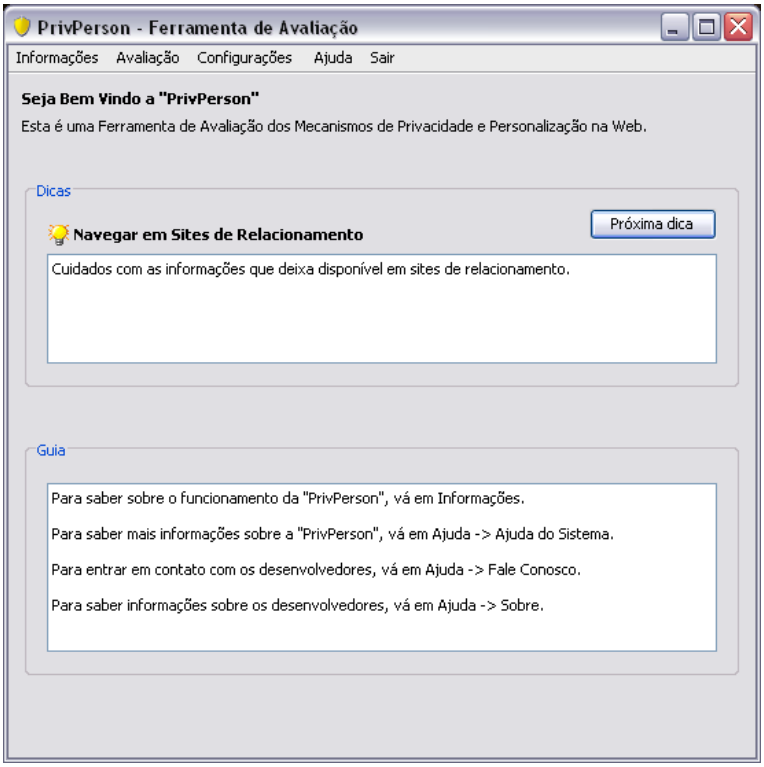


**Figura 21. Tela principal da "PrivPerson"**

Nesta figura, além do menu das principais funcionalidades disponibilizadas pela “PrivPerson” são também apresentadas outras informações, como saudações e dicas, às quais têm o intuito de alertar os usuários sobre assuntos relevantes à navegação na web, e guias para instruí-los na utilização da “PrivPerson”.

Através do menu de funcionalidades são disponibilizadas informações aos usuários de como utilizar a “PrivPerson”; características apoiadas e os passos seguidos para seu desenvolvimento; propósito de sua utilização; taxonomias de privacidade e personalização utilizadas; como foi feita a quantificação das camadas, seguidas de seus respectivos pesos; opções de avaliação dos mecanismos; ajuda; contatos e desenvolvedores, bem como a instituição que apoiou o desenvolvimento.

As informações estáticas utilizadas pela “PrivPerson” foram armazenadas em arquivos para a persistência dos dados, sejam elas informações disponibilizadas aos usuários ou que serviram de base para funcionamento correto da avaliação dos mecanismos.

Esses arquivos estão no formato XML, sendo recuperados para fins de análises dos mecanismos durante a avaliação, por motivos de atualizações e para disponibilizar informações aos usuários. As informações são recuperadas e exibidas aos usuários, sendo algumas delas: nome das ferramentas e dos cenários de utilização, descrição, camadas constituintes das taxonomias e seus respectivos pesos.

Em relação às dicas, essas são informações sobre assuntos referentes à privacidade, à personalização e à segurança dos usuários, tendo sido cadastradas e armazenadas em arquivos XML, seguindo os padrões da linguagem. A partir desses arquivos, essas são recuperadas e expostas na tela inicial aleatoriamente, podendo os usuários navegar por entre elas através do botão “Próxima dica”.

Na Figura 22 é mostrada a tela inicial da “PrivPerson” com o menu Informações selecionado, o qual aborda assuntos gerais sobre a “PrivPerson”, sobre as taxonomias utilizadas na avaliação dos mecanismos, com detalhamento dos pesos atribuídos para cada camada.

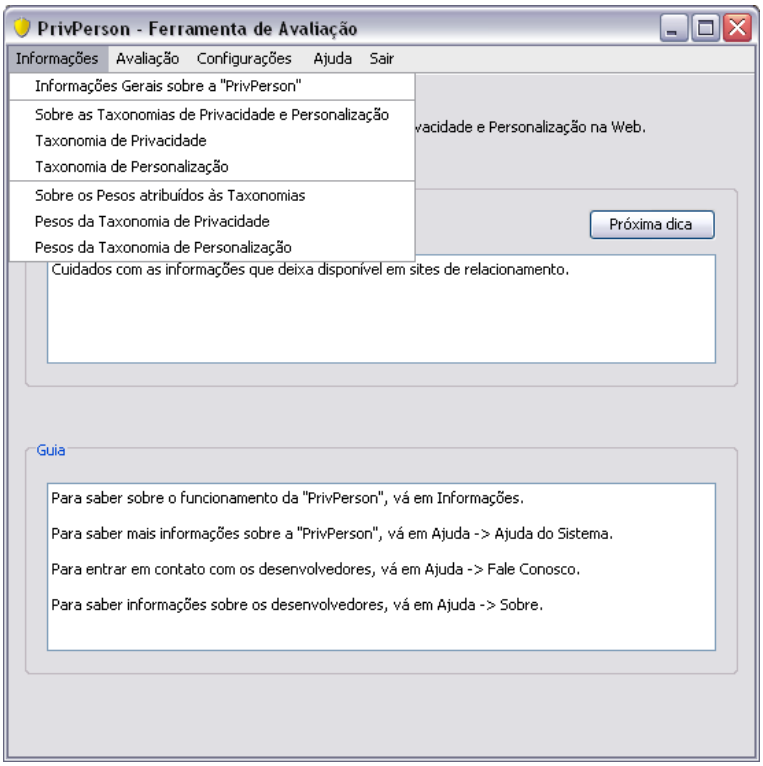


**Figura 22. Menu Informações**

Nesta tela, apresentada na Figura 22, são especificadas as características utilizadas para o funcionamento da "PrivPerson", sendo elas: i) informações gerais sobre a "PrivPerson", apresentado os objetivos para seu desenvolvimento; ii) descrição das taxonomias de privacidade e personalização utilizadas, enfatizando o porquê da utilização dessas; iii) metodologia seguida para atribuição de pesos a cada uma das camadas das taxonomias apresentadas e, iv) classificação de relevância apresentada pelas camadas e os pesos atribuídos a cada uma.

Para detalhar ainda mais a explicação, na "PrivPerson" também é disponibilizada uma imagem do funcionamento da avaliação dos mecanismos, retratando a análise e os resultados disponibilizados pela avaliação, o que facilita o entendimento dos usuários.

Na tela onde são disponibilizadas informações de ajuda aos usuários, são descritos os passos que devem ser seguidos para que seja feita a avaliação das ferramentas, de sites e de cenários de utilização, individualmente.

Não foram mostradas todas as telas da "PrivPerson" devido a algumas terem o objetivo de apenas detalhar o funcionamento e as técnicas utilizadas para seu desenvolvimento, sendo esses temas descritos nos capítulos apresentados e ao longo deste.

## 6.2. Configurações da “PrivPerson”

A “PrivPerson” possibilita a análise dos mecanismos de forma quantificada, disponibilizando informações relevantes aos usuários e oferecendo ainda ambientes distintos aos usuários e administradores, sendo uma ferramenta com interfaces amigáveis.

A configuração da “PrivPerson” pode ser feita pelo menu Configurações, disponível apenas ao administrador. Nesse módulo é necessária uma senha de administração, a qual assegura que usuários comuns não poderão ter acesso a configurações internas, de modo a impedir que informações erradas sejam cadastradas na “PrivPerson”.

Dessa forma, são disponibilizadas telas para gerenciamento das ferramentas e dos cenários de utilização. Nas telas podem ser visualizadas todas as ferramentas e cenários já cadastrados, seguidos de suas informações e das camadas das taxonomias de privacidade e personalização que contemplam.

A possibilidade de gerenciamento foi implementada de forma que a “PrivPerson” possa ser atualizada sempre que preciso, devido à sua característica inerente de ser extensível, não sendo suas análises limitadas a apenas as ferramentas e cenários estudados durante seu desenvolvimento.

As opções de cadastro, edição e exclusão das ferramentas e cenários, bem como ações para salvar, próximo e anterior, podem ser executadas pelos administradores na “PrivPerson” a partir do menu de navegação. Esse contém ícones representando as opções e ações que podem ser executadas, de modo que a interface seja amigável.

As telas de gerenciamento encontram-se preenchidas com informações de ferramentas ou cenários já cadastrados, podendo todas serem visualizadas através do menu de navegação. Quando o usuário clica em Cadastrar Ferramenta, na tela de gerenciamento de ferramentas, todos os campos da tela são apagados para que informações da nova ferramenta possam ser inseridas. Para efetivar o cadastro é necessário clicar no ícone Salvar Ferramenta, como mostrado na Figura 23.

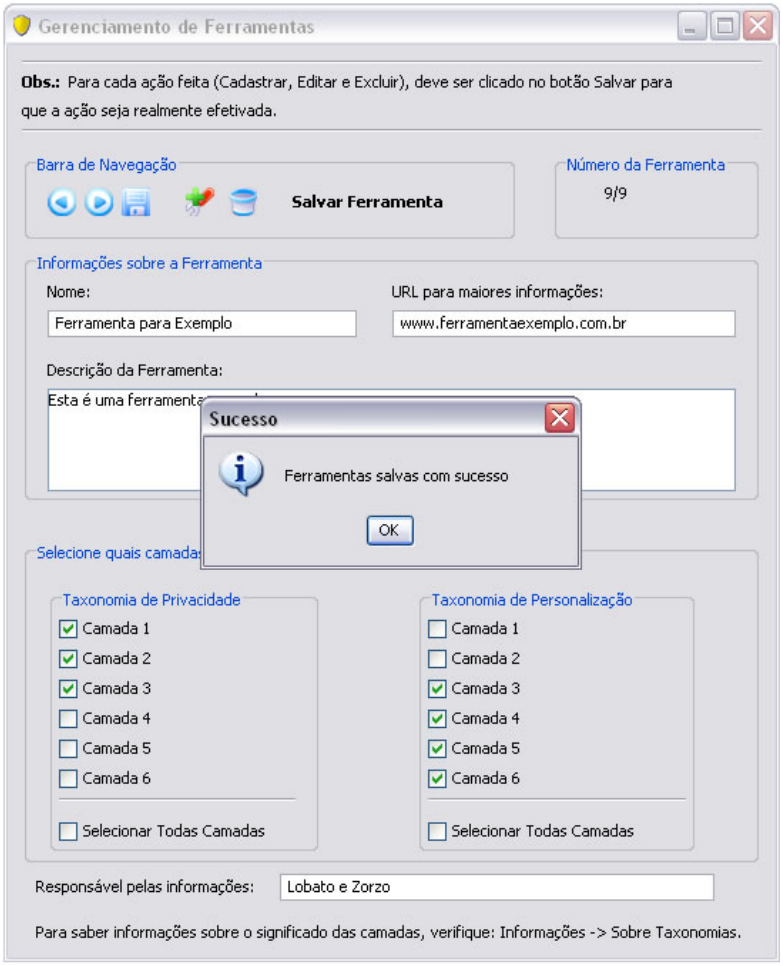


**Figura 23. Cadastro de nova Ferramenta**

Nesta figura ainda são mostradas informações como o número total de ferramentas já cadastradas na "PrivPerson", o número da posição de cadastro da ferramenta que está sendo exibida, informações referentes a ela e as camadas contempladas, sendo essas marcadas quando forem utilizadas pela ferramenta.

Na exclusão das ferramentas ou cenários de utilização, para que a ação possa ser efetivada antes da exclusão é mostrada uma janela de aviso, perguntando se o usuário tem certeza que deseja excluir tais informações, e que após a exclusão tal ação não poderá ser desfeita. Após a exclusão de uma ferramenta ou cenário de utilização, o número total referente ao mecanismo é decrescido de 1 e então é exibido ao usuário o mecanismo que antecede ao que foi excluído.

Várias medidas de segurança foram tomadas para evitar que aconteçam erros durante a manipulação dos mecanismos, como por exemplo: caso sejam feitas alterações pelo administrador e essas não sejam salvas, ao fechar a janela é exibida uma mensagem, sob a forma de alerta, perguntando se deseja salvar as alterações feitas.

As mesmas ações e medidas de segurança apresentadas na tela de configuração das ferramentas estão também disponíveis para os cenários de utilização, sendo mostrado na Figura 24.

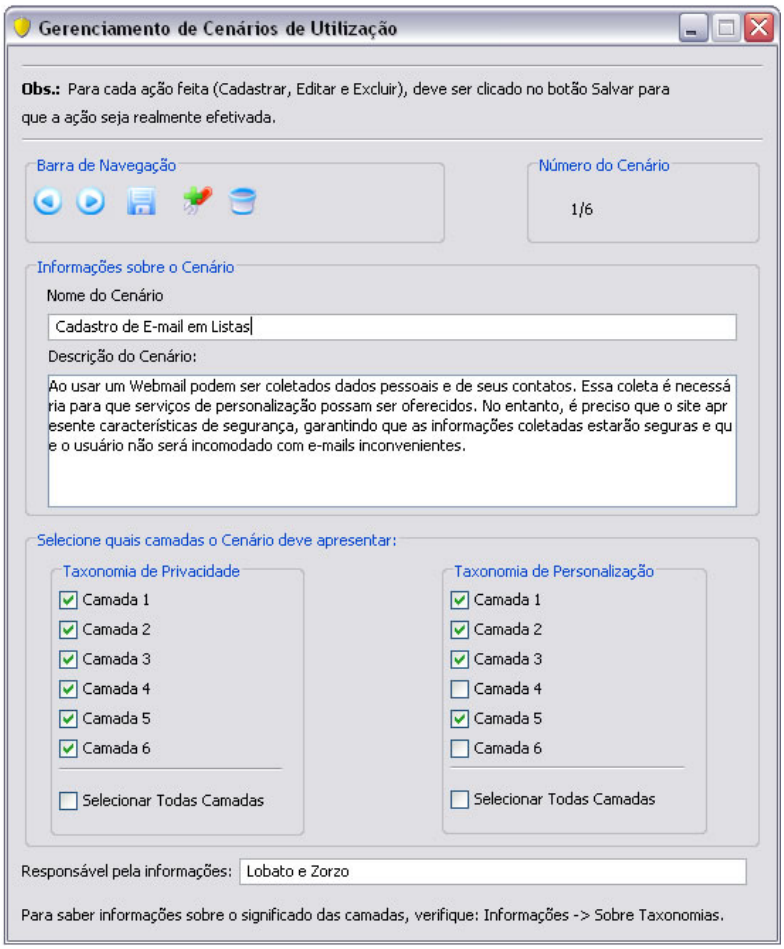


**Figura 24. Tela de gerenciamento dos Cenários de Utilização**

Não serão detalhadas as funcionalidades dos cenários de utilização, de modo a evitar redundância de informações, já que são as mesmas apresentadas pelas ferramentas, seguindo um mesmo padrão e lógica.

No momento do cadastro de novas ferramentas e cenários de utilização, um campo é apresentado na tela para que seja informado o nome da pessoa responsável pelo cadastramento. Isso é feito para permitir que seja identificado o responsável pelas informações cadastradas, de modo a garantir a veracidade dessas.

A seguir são mostradas as telas da "PrivPerson" para avaliação dos mecanismos: ferramentas, sites e cenários de utilização. Essas são funcionalidades mais relevantes da "PrivPerson", pois serviram para validar de forma prática a utilização das taxonomias de privacidade e personalização.

## 6.3. Avaliação dos Mecanismos

Depois de já terem sido apresentados os passos seguidos para o desenvolvimento da "PrivPerson" e algumas de suas funcionalidades, nesta seção é apresentada a construção do módulo de avaliação dos mecanismos.

Através da "PrivPerson" tornou-se possível retornar aos usuários, de forma mensurável, os níveis de privacidade e de personalização oferecido pelos mecanismos, verificando-se quais das camadas esses contemplam.

As camadas foram parametrizadas com a utilização de pesos, mostrados na seção 5.4, de forma que, sabendo-se quais camadas o mecanismo em avaliação contempla, é possível quantificar a privacidade e a personalização oferecidas aos usuários.

Utiliza-se valores para mensurar o quão seguros e eficientes os mecanismos são, pois através de valores reais é mais fácil para o usuário entender o nível de privacidade e personalização oferecida.

Esse nível é apresentado aos usuários na forma de porcentagem para facilitar o entendimento do usuário e, conhecido o resultado da avaliação, tornou-se possível aos usuários a utilização dos mecanismos de forma consciente.

Na Figura 25 é mostrada a tela principal da "PrivPerson", com a opção do usuário escolher qual mecanismo deseja que seja avaliado através do menu Avaliação.

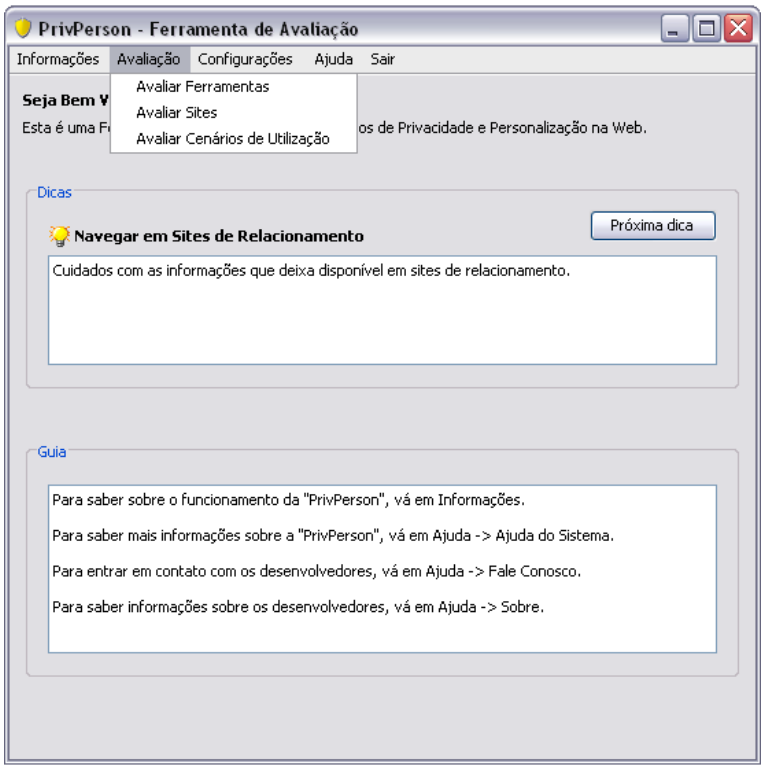


**Figura 25. Menu para avaliação dos mecanismos**

Para fazer a avaliação do nível de privacidade e personalização apresentado pelas ferramentas e cenários de utilização, recupera-se as informações armazenadas nos arquivos XMLs sobre quais camadas esses contemplam.

A avaliação dos sites feita pela "PrivPerson", foi disposta de modo que se pudesse comparar com a análise manual apresentada no Estudo de Caso, apresentado no APÊNDICE A. Dessa forma foi possível chegar a resultados do quão eficiente a "PrivPerson" é, validando sua análise automatizada com a análise manual.

A seguir é apresentada, de forma detalhada a avaliação feita em cada um dos mecanismos, sendo elas: Avaliar Ferramentas, Avaliar Sites e Avaliar Cenários de Utilização.

### 6.3.1. Avaliação das Ferramentas

Na Figura 26 é mostrado o funcionamento do módulo Avaliar Ferramentas, exemplificando a maneira como a "PrivPerson" foi implementada para avaliação das ferramentas que garantem a privacidade dos usuários e oferecem serviços personalizados.

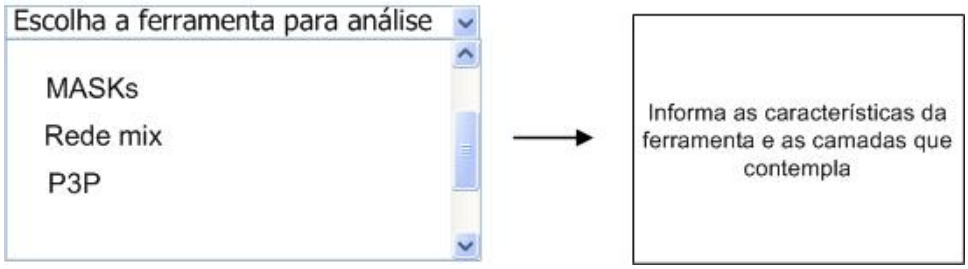


**Figura 26. Exemplificação da avaliação das ferramentas**

Pela "PrivPerson" são apresentados ao usuário opções de ferramentas previamente cadastradas. De acordo com a escolha do usuário é retornada a análise da ferramenta, descrevendo o nível de privacidade e de personalização oferecido pela mesma.

As ferramentas estudadas, cujas informações sobre suas funcionalidades foram armazenadas na "PrivPerson", se enquadram em algumas das camadas das taxonomias de privacidade e de personalização, como mostrado na Tabela 5.

**Tabela 5. Associação entre ferramentas e camadas**

| Ferramenta estudada | Camadas de Privacidade | Camadas de Personalização |
|---|---|---|
| 1 - MASKs | Camada 4, 5, 6 | Camada 1, 2, 3, 4, 5 |
| 2 - Pseudônimo | Camada 4, 6 | Camada 2, 3, 4, 5 |
| 3 – Mix Kobsa | Camada 6 | Camada 2, 3, 4, 5 |
| 4 - Onion | Camada 6 | Camada 2, 4, 6 |
| 5 - Crowd | Camada 6 | Camada 2, 4, 6 |
| 6 – P3P | Camada 2, 4, 6 | Camada 3, 5 |

A partir do conhecimento de quais camadas as ferramentas contemplam, tornou-se possível retornar ao usuário o nível de privacidade e personalização oferecido, visto que os pesos foram necessários para a definição desse nível.

A utilização das respostas na forma de nível foi importante para que os usuários pudessem ter um valor estimado sobre o que é oferecido pelos mecanismos, sendo esses valores apresentados sob a forma de porcentagem, que varia de 0% a 100%.

Durante a análise é mostrada uma barra de status para retratar o andamento da análise e ao final, essa barra é preenchida de acordo com o nível, ou porcentagem, de privacidade e personalização disponibilizado.

Na Figura 27 é mostrada a tela de avaliação das ferramentas cadastradas na "PrivPerson", tomando como exemplo o MASKS.

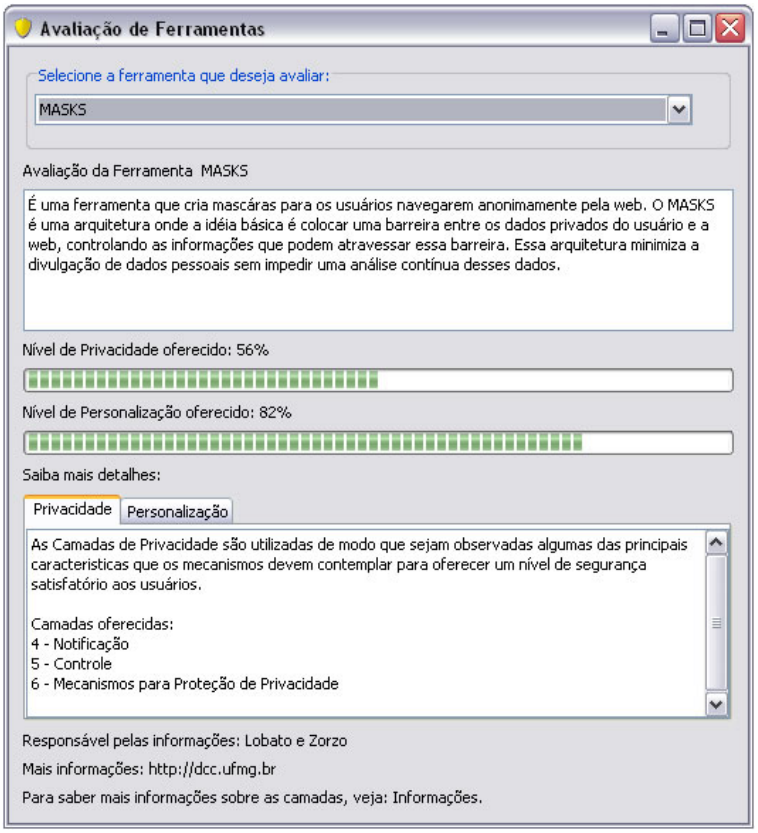


**Figura 27. Tela da "PrivPerson" para avaliação de ferramentas**

Pode-se observar que o MASKS apresenta 3 das camadas de privacidade, sendo elas: Camada 4 - Notificação; Camada 5 - Controle e Camada 6 - Mecanismo para Proteção de Privacidade, oferecendo 56% de privacidade. O MASKS ainda aborda 5 das camadas da taxonomia de personalização, sendo: Camada 1 - Controle de Filtros; Camada 2 - Coleta de Dados; Camada 3 - Transferência de Dados; Camada 4 - Utilização de Intermediário e Camada 5 - Armazenamento em Servidor, apresentando 82% de personalização.

Tendo conhecimento da porcentagem de privacidade e personalização oferecida, o usuário pode tomar decisões estando mais seguro quanto a quais ferramentas utilizar.

A partir do resultado da avaliação da ferramenta ou do cenário de utilização, o usuário pode ter mais detalhes, em "Mais Informações", bem como quais camadas são contempladas e o nome do responsável pelo cadastro da ferramenta na "PrivPerson".

### 6.3.2. Avaliação dos Sites

Nesta seção é detalhado o módulo de análise dos sites, abordando as etapas seguidas para a implementação e ainda, a importância do *crawler*[23] desenvolvido para fazer análises nos códigos dos sites.

Para a avaliação dos sites não existem opções de cadastros, edição e remoção, pois esses devem ser informados pelos usuários no momento da análise e a avaliação desses deve ser feita em tempo real, sendo mostrados aos usuários o *status* da análise.

Na Figura 28 é mostrada uma exemplificação da implementação da avaliação dos sites, onde é feita a procura pela *tag form*, uma das chaves para a verificação da utilização da Camada 2 – Coleta de Dados da taxonomia de personalização.

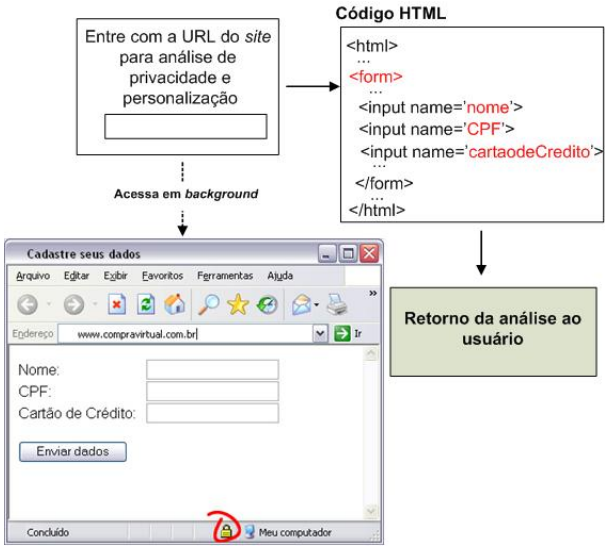


**Figura 28. Exemplificação da análise de site**

A partir de uma URL informada pelo usuário o *crawler* da "PrivPerson" acessa as páginas referentes ao site de maneira automatizada, desde que essa URL seja válida e acessível.

[23] Um *crawler* ou *robot* é um programa que pesquisa sistematicamente a Internet à procura de informações para indexação. Em geral, é um programa que segue todos os *links* existentes numa página web, para encontrar novas páginas e novos *links* sucessivamente (MOTA e GOMES, 2004).

Para que a avaliação do site possa ser realizada de maneira satisfatória, o *crawler* faz *download* das páginas referenciadas pelo site e, a partir daí, analisa o código fonte à procura de características, sendo essas, palavras ou *tags*, que identifiquem a utilização das camadas das taxonomias de privacidade e personalização. O acesso às páginas dos sites acontece de forma transparente ao usuário.

Durante o *download* das páginas algumas restrições impostas na implementação são efetuadas. Um exemplo é a verificação do tipo do arquivo para que não seja baixado tudo o que o site referenciar, de modo que arquivos com extensões suspeitas, como .exe, e arquivos que não referenciam páginas web, como .pdf e .jpg, não sejam analisadas durante a avaliação do site, otimizando dessa forma a busca do *crawler*.

Para cada camada a ser analisada nos sites, existem palavras que são relevantes à sua existência, sendo necessário um analisador de regras gramaticais (*parsing*) para efetuar a análise do código.

Existem camadas nas taxonomias cujas utilizações devem ser verificadas através de *tags* padronizadas computacionalmente. Como por exemplo, o uso de *cookies*, é verificado pelo cabeçalho da requisição HTTP e a utilização de ambiente seguro HTTPS, é verificado pelo protocolo de transmissão dos arquivos.

Utiliza-se ainda para verificar se os sites apresentam as camadas, expressões regulares próprias, para o encontro (*matching*) de palavras e expressões que contemplam as taxonomias.

Com o desenvolvimento do Estudo de Caso, APÊNDICE A, foi possível verificar como os sites disponibilizam as informações, as quais são referentes a algumas das camadas das taxonomias. Dessa forma, foi possível observar algumas palavras-chaves e abastecer o *crawler* com essas, de modo que seja possível reconhecer a utilização dessas pelos sites e conseqüentemente, reconhecer a utilização das camadas das taxonomias.

Foi necessária a observação dessas palavras-chaves, pois diferentes sites podem apresentar diferentes maneiras de tratar os elementos e os serviços disponibilizados, dificultando dessa forma o encontro desses. Por exemplo, a Política de Privacidade apresentada pelos sites, onde grande parte das informações relevantes são mostradas, pode ser disponibilizada das mais diferentes maneiras.

A não padronização de certas informações pode dificultar a análise dessas pelos usuários e pelas ferramentas automatizadas. No entanto, para suprir esse problema, é apresentada uma padronização para as Políticas de Privacidade, conforme abordado no APÊNDICE B, a qual foi desenvolvida utilizando algumas das camadas das taxonomias de privacidade e personalização.

A busca feita pelo *crawler* pelas páginas que são referenciadas no site é feita em ramificações, iniciando na página raiz e descendo o nível de avaliação pelos demais *links*. Podem ser avaliados todos os *links* referenciados pelo site dentro ou não do domínio do site e até o número máximo de URL's desejadas.

A Figura 29 mostra a tela da "PrivPerson" para a avaliação dos sites da web.

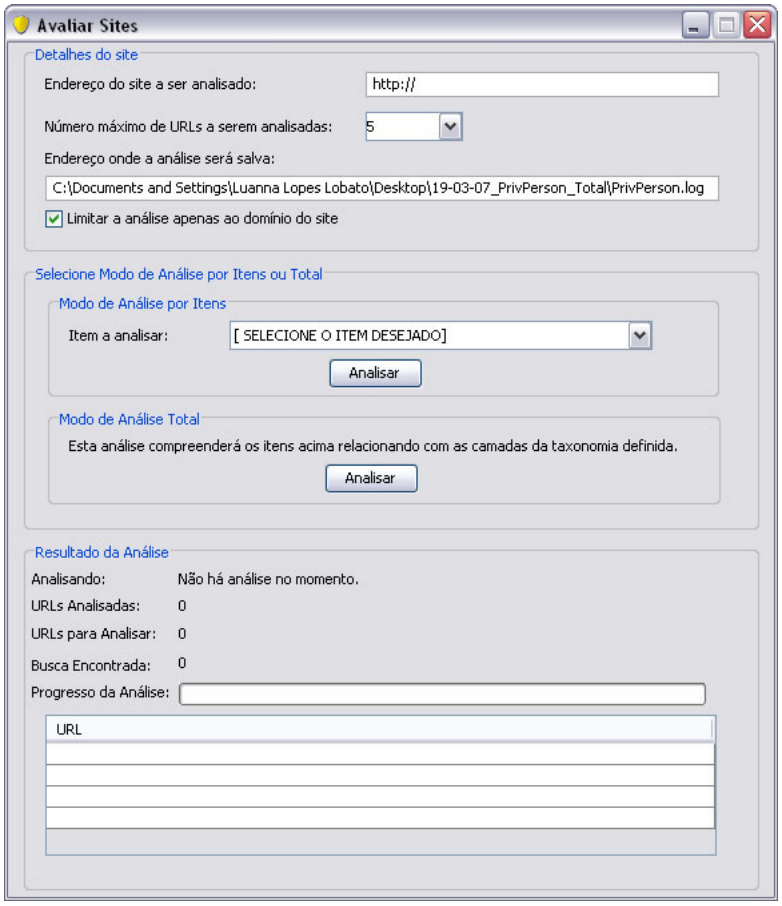


**Figura 29. Tela da "PrivPerson" para avaliação dos sites**

Pode-se verificar que é exibido ao usuário um campo onde o mesmo pode informar a URL do site a ser avaliado. Se a URL não estiver acessível, é então retornada ao usuário, em forma de caixa de diálogo, uma informação avisando o ocorrido.

No entanto, tal avaliação pode ser demorada e cansativa quando um site tem referência a muitas páginas. Para solucionar esse problema há a possibilidade do usuário limitar essa análise ao número de páginas que deseja que a "PrivPerson" avalie e ao domínio do site.

Apesar dessa ser uma solução para o tamanho da análise, pode apresentar desvantagens se o usuário limitar sua pesquisa em poucas buscas, pois apenas com poucas URLs a serem analisadas pode ser que o *crawler* não encontre nos sites informações referentes às camadas. Dessa forma, uma solução é limitar a análise a apenas o domínio do

site e, se esse referenciar muitas outras páginas, limitar o tamanho da análise a um número considerável, como mostrado na Figura 30.

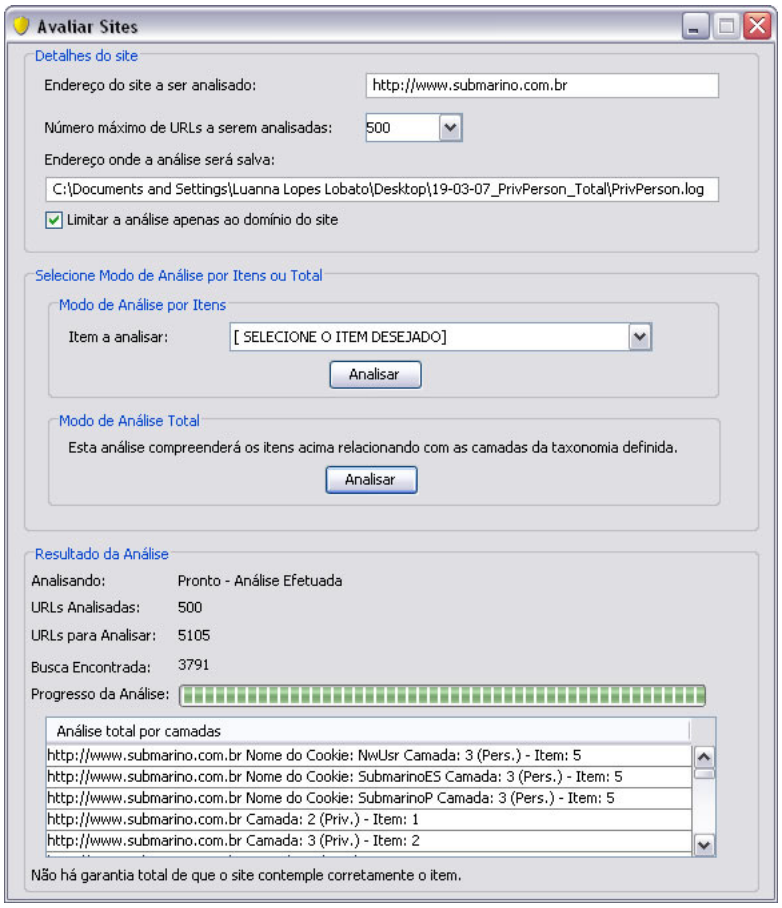


**Figura 30. Análise do site limitada**

Nessa figura pode ser verificado que o resultado da análise é informado ao usuário, mostrando em quais URLs as camadas foram encontradas, tendo a análise sido restrita apenas ao domínio do site.

Se o número máximo de URLs a serem avaliadas, determinadas pelo usuário, for maior que o número de páginas pertencentes ao domínio do site, não serão analisadas todas, para que não infrinja a limitação de páginas ao domínio. Dessa forma, é possível avaliar de forma significativa a privacidade e personalização que está sendo oferecida ao usuário pelo site, pois o *crawler* não descerá a pesquisa por *links* que fazem referência a outros sites.

A avaliação feita nos sites foi disposta em módulos: Modo de Análise por Itens e Modo de Análise Total. No primeiro, é analisado se os itens, que representam as camadas, são apresentados pelos sites, podendo o usuário selecionar um item de cada vez. No segundo, Modo de Análise Total, os sites são analisados à procura de todas as características que identifiquem o uso das camadas das taxonomias de privacidade e de personalização.

Os dois modos de análise são relevantes. A análise dos sites por itens é importante pois o usuário pode desejar verificar se o site contempla uma característica específica, não sendo necessário esperar uma análise completa abrangendo a verificação da ocorrência de todas as camadas do site. Também é suficiente pois com o uso dos itens é possível ao usuário identificar o que deseja analisar nos sites, sendo retornado a ele se o site apresenta ou não o item verificado.

No entanto, a análise total apresenta-se mais eficiente quando deseja-se verificar o nível de privacidade e personalização oferecido, já que nessa, para cada página analisada é verificada a utilização de todas as camadas das taxonomias. Dessa forma, otimiza-se as buscas e é retornado ao final a porcentagem de privacidade e personalização oferecida pelos sites.

No Modo de Análise por Itens são disponibilizados 15 itens, apresentados no APÊNDICE A, os quais são descrições detalhadas das camadas referentes as taxonomias de privacidade e personalização. Esses itens foram implementados e disponibilizados em uma caixa para seleção, de forma a permitir 15 módulos diferentes de análise nos sites, sendo a análise por um dos itens mostrada na Figura 31.

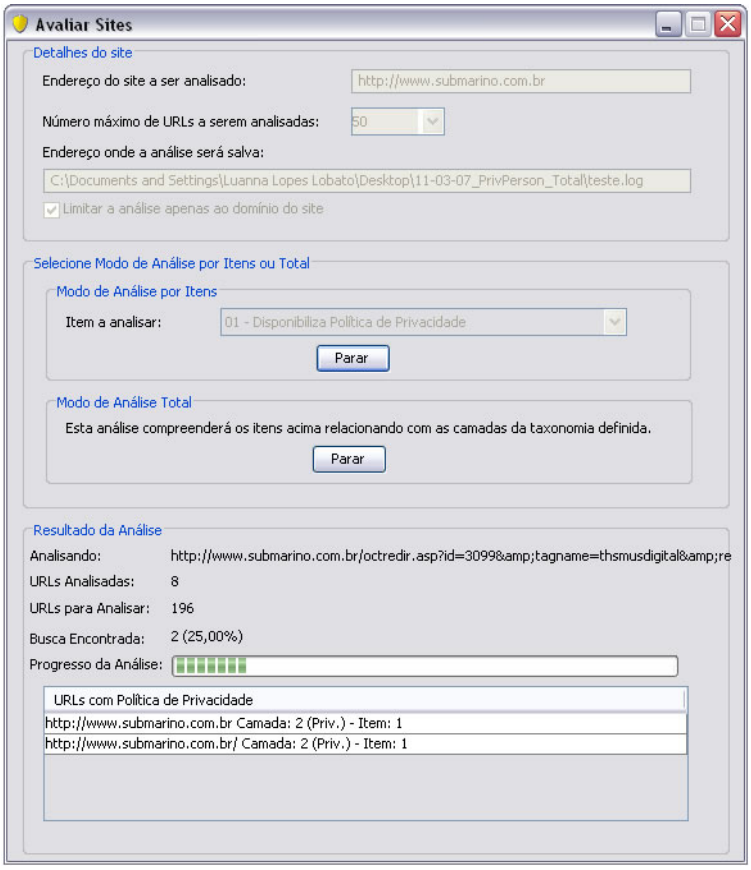


**Figura 31. Análise do site baseado em itens**

Na Figura 31 é apresentada a análise de um site, baseada em item, verificando se esse apresenta o Item 01 - Disponibiliza Política de Privacidade, o qual representa a Camada 2 - Políticas de Privacidade da taxonomia de privacidade.

Observa-se que durante a avaliação os campos ficam desabilitados, não sendo possível clicá-los. Apenas é possível clicar no botão "Parar", que no início da análise tinha o nome "Pesquisar", o qual encontra-se habilitado e nomeado como "Parar" até o fim da análise ou até que a análise seja cancelada.

No Modo de Análise Total o *crawler* verifica o código fonte das páginas referenciadas no site em busca de comprovações sobre a utilização de todas as camadas das taxonomias, retornando as camadas utilizadas pelo site e o nível de privacidade e personalização oferecido ao usuário. Na Figura 32 é mostrado o resultado de uma avaliação total.

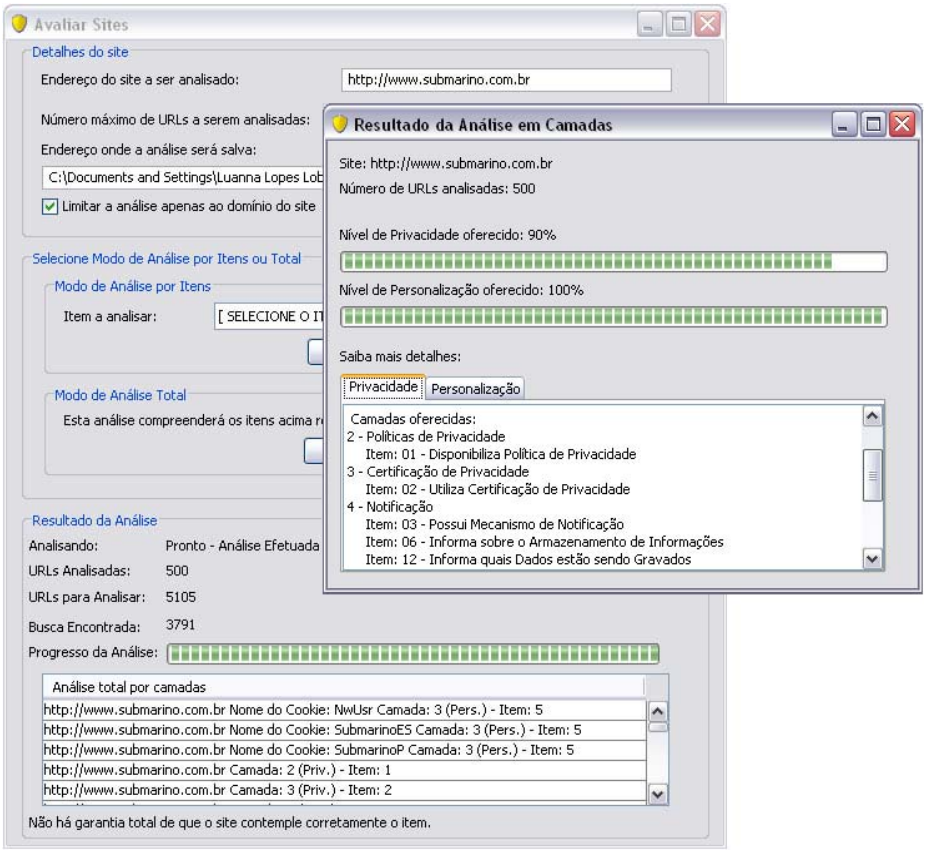


**Figura 32. Análise no site de todas as camadas das taxonomias**

Para otimização da análise dos sites são utilizadas *threads*, as quais permitem execuções paralelas da análise de diferentes páginas em um mesmo espaço de tempo.

Durante os dois tipos de análise informações sobre a avaliação vão sendo disponibilizadas, comunicando seu andamento. É informado o número de URLs já analisadas; a serem analisadas; o número de buscas encontradas; a barra de status da análise; e, as URLs que contêm a análise feita, onde tais camadas e itens foram encontradas pelo *crawler*.

Um detalhe importante feito durante o desenvolvimento do *crawler* é que, ao analisar o código de uma página, outros *links* são reconhecidos, sendo esses inseridos em uma tabela *hash* formando a árvore de busca na qual a análise será feita. Dessa forma, é evitada a análise repetitiva dos *links*, o que poderia resultar em um *loop* infinito.

A "PrivPerson" gera um arquivo de *log* para o armazenamento da resposta da análise dos sites, no entanto tal arquivo é substituído a cada nova análise, pois o nome do *log* e o endereço do caminho a ser armazenado são os mesmos.

Caso o usuário queira guardar automaticamente o resultado das análises, esse pode definir um novo nome a ser atribuído ao arquivo de *log* a cada nova análise, ou então informar um novo caminho onde o *log* deva ser armazenado. Essa possibilidade de alteração de nome e caminho encontra-se na tela de avaliação dos sites, a qual já traz um caminho e nome de *log* padrão.

### 6.3.3. Avaliação dos Cenários de Utilização

Para a escolha de quais seriam os cenários de utilização a serem avaliados, utilizou-se a metodologia de ações mais comuns efetuadas pelos usuários na web. Essas ações foram disponibilizadas na "PrivPerson" em uma caixa de seleção, onde o usuário poderá escolher qual cenário deseja avaliar. A Figura 33 mostra uma exemplificação do funcionamento da avaliação dos cenários de utilização.

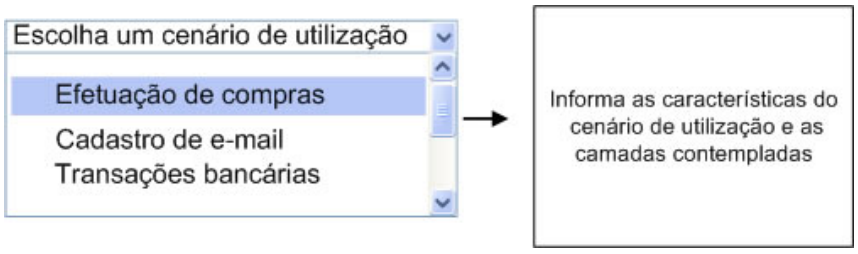


**Figura 33. Exemplificação da análise de cenário de utilização**

Nessa figura pode-se visualizar que na resposta da avaliação do cenário, retornada ao usuário, são também identificadas quais camadas devem estar presentes no ambiente.

A seguir, na Tabela 6, são apresentados os cenários de utilização já cadastrados na "PrivPerson", seguidos das camadas de privacidade e personalização que esses deveriam contemplar.

**Tabela 6. Associação entre cenários de utilização e camadas**

| Cenários | Camadas de Privacidade | Camadas de Personalização |
|---|---|---|
| 1 – Cadastro de e-mail em listas | Camada 1, 2, 3, 4, 5, 6 | Camada 1, 2, 3, 5 |
| 2 – Compra on line | Camada 1, 2, 3, 4, 5, 6 | Camada 1, 2, 3, 4, 5, 6 |

| | | |
|---|---|---|
| 3 – Navegar em sites de relacionamento | Camada 1, 2, 3, 4, 5, 6 | Camada 1, 2, 3, 5 |
| 4 – Efetuar transações bancárias | Camada 1, 2, 3, 4, 5, 6 | Camada 1, 2, 3, 4, 5, 6 |
| 5 – Cadastro em sites | Camada 1, 2, 3, 4, 5, 6 | Camada 1, 2, 3, 4, 5, 6 |

A partir da escolha de qual cenário de utilização o usuário deseja visualizar a análise, é retornado a ele, de forma quantificada, o nível de privacidade e personalização ideal que o cenário deve atingir, seguindo de informações aos usuários, como mostrado na tela de avaliação de cenários, Figura 34.

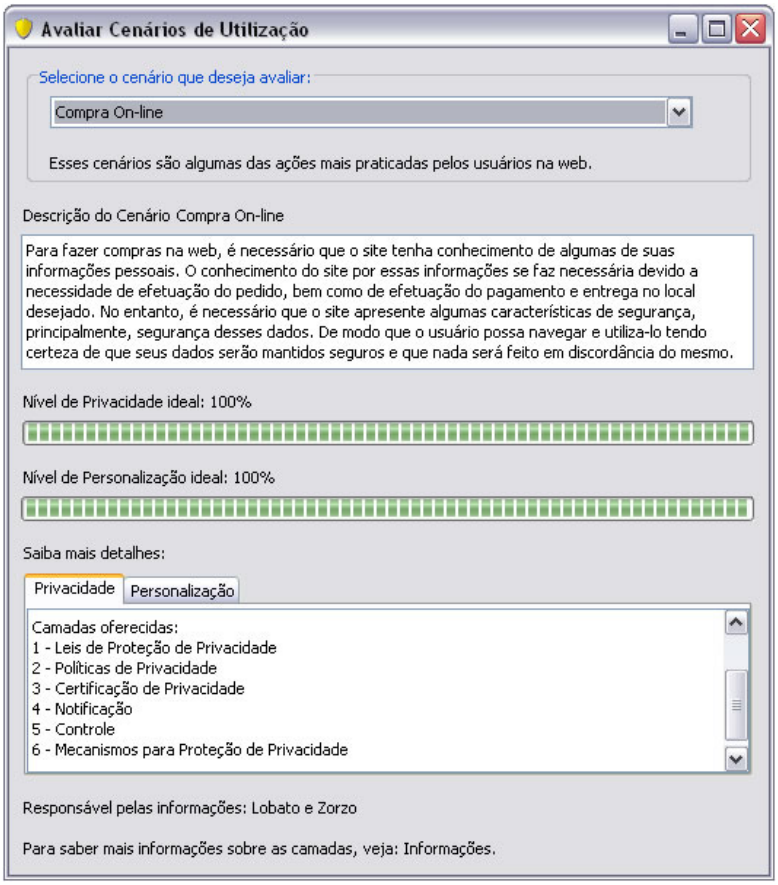


**Figura 34. Tela da "PrivPerson" para avaliação dos cenários de utilização**

Nessa tela, é possível observar que são retornados ao usuário informações sobre a privacidade e a personalização, divididas por abas de acesso. Nessas abas os usuários poderão verificar quais camadas os cenários de utilização deverão apresentar para possibilitar a segurança e ainda oferecer serviços de personalização aos usuários. Ainda apresentam informações sobre a privacidade e personalização oferecidas aos mesmos.

O nível de privacidade e de personalização foi colocado como o nível ideal que o cenário de utilização deve apresentar, pois esses podem ser disponibilizados aos usuários de várias formas, dependendo da empresa que os implementam e das regras a serem seguidas. Dessa forma, não foi possível analisá-los como as ferramentas foram analisadas, pois cada um pode representar uma abordagem diferente.

## 6.4. Avaliação e Validação da “PrivPerson”

Sendo a usabilidade e a satisfação do usuário duas das características mais importantes que devem ser apresentadas para a qualidade de software, foram aplicadas, neste trabalho, técnicas de modo a medi-las. Isso foi feito pois, entende-se que o desenvolvimento de software com vistas à satisfação do usuário produzirá melhor qualidade.

Após ser apresenta a “PrivPerson”, a seguir são retratados os procedimentos seguidos para realização de um estudo de usabilidade, envolvendo seres humanos, para avaliação da eficiência e da relevância da mesma. Nesta seção, também é mostrada a metodologia aplicada para seleção dos usuários a participarem dessa pesquisa.

Os resultados deste estudo são apresentados no próximo capítulo, objetivando medir a satisfação apresentada pelos participantes na utilização da “PrivPerson”, com vistas a aumentar a eficiência e a relevância da ferramenta.

Para realização do estudo de caso seguiram-se os requisitos da Resolução CONEP 196/96, de modo a respeitar todas as restrições impostas necessárias para o desenvolvimento de um estudo com seres humanos, sendo descritos a seguir. Tal projeto desenvolvido para aplicação do estudo, foi aprovado tendo o número de protocolo 047/2007 (LOBATO e ZORZO, 2007d).

### 6.4.1. Planejamento

Com a utilização da “PrivPerson” para avaliação dos mecanismos e conseqüentemente com os resultados dessa avaliação, os participantes puderam mensurar o quanto de privacidade e personalização é oferecido, a eficiência da “PrivPerson”, a precisão das respostas e a real contribuição em seu desenvolvimento.

A satisfação encontrada pelos participantes da amostra foi avaliada através de testes de execução baseados em um questionário auto-explicativo. Esses testes foram aplicados após a fase do projeto de implementação da “PrivPerson” e os resultados foram documentados e analisados, sendo esses referências à utilização e aos testes da mesma.

O local escolhido para a execução do estudo foi o laboratório Aquário, pertencente ao Grupo de Sistemas Distribuídos e Redes do Departamento de Computação da UFSCar, São Carlos, SP.

Após a definição do local, a próxima etapa envolveu a busca e a definição de uma amostra constituída por participantes, os quais contribuíram para a execução do estudo empregado.

Para se determinar o tipo e o tamanho da amostra várias questões foram levadas em consideração, como a familiarização com a navegação web, o conhecimento sobre compras on-line e, principalmente, a facilidade de contato com os participantes.

Segundo as recomendações de Holzinger (2005) é interessante avaliar 20 ou mais usuários quando se realiza testes de usabilidade. Assim, foram tomados como amostra 20 participantes, aqui também chamados de usuários, sendo esses alunos da graduação, pós-graduação e professores do Departamento de Computação da UFSCar.

### 6.4.2. Termo de Consentimento

Para que pudesse ser realizado o teste de análise da "PrivPerson" com usuários foi necessário seguir os parâmetros impostos pela comissão de ética da UFSCar.

Uma das regras impostas foi o desenvolvimento e apresentação de um termo de consentimento aos participantes, o qual comprova a existência e consentimento deles sobre a contribuição no estudo, sendo esse termo apresentado na Tabela 7.

**Tabela 7. Termo de Consentimento Livre e Esclarecido**

| **Termo de Consentimento Livre e Esclarecido** |
|---|
| 1) Você foi selecionado por sua formação profissional, sua familiaridade na navegação pela web, facilidade de acesso no local onde serão realizados os experimentos, no entanto sua participação não é obrigatória. |
| 2) A qualquer momento você pode desistir de participar e retirar seu consentimento. |
| 3) Sua recusa não trará nenhum prejuízo em sua relação com os pesquisadores, responsáveis por este estudo, Luanna Lopes Lobato e Prof. Dr. Sérgio Donizetti Zorzo, ou com a instituição onde tal estudo é aplicado, sendo essa o Departamento de Computação da UFSCar. |
| 4) Os objetivos deste estudo estão embasados na possibilidade da análise qualitativa da ferramenta, "PrivPerson", desenvolvida para avaliar a privacidade e personalização oferecida pelos mecanismos (ferramentas, sites e cenários de utilização) aos usuários, os qual objetiva o consentimento do usuário ao utilizar tais mecanismos, visando a garantia de sua privacidade, bem como a oferta de serviços personalizados. |
| 5) Sua participação nesta pesquisa consistirá na utilização da "PrivPerson", podendo avaliar as ferramentas, cenários de utilização já cadastrados na "PrivPerson", sites de comércio eletrônico da web e na resposta a um questionário, o qual objetiva coletar informações referentes a satisfação na utilização da mesma. |
| 6) Os benefícios relacionados com a sua participação estão ligados à contribuição para a avaliação do trabalho de mestrado proposto e um ganho, tendo esse como sendo um material que poderá ser referenciado no ambiente de pesquisa. |
| 7) As informações obtidas através dessa pesquisa serão confidencias, assegurando que o sigilo sobre sua participação e sobre informações pessoais, como seu nome, serão preservados. |
| 8) Os dados não serão divulgados de forma a possibilitar sua identificação. As informações |

coletadas não estarão vinculadas à sua identidade.

9) Você receberá uma cópia deste termo onde consta o telefone e o endereço de onde encontrar a pesquisadora principal, Luanna Lopes Lobato, podendo tirar suas dúvidas sobre o projeto e sua participação, agora ou a qualquer momento após a participação no estudo.

_____
Luanna Lopes Lobato
Email: luannalopeslobato@gmail.com

Universidade Federal de São Carlos (UFSCar) - Departamento de Computação
Rodovia Washington Luis, km 235, Cep: 13565-905 São Carlos – SP
Telefone: (16) 3351-8572

**Declaro que entendi os objetivos, riscos e benefícios de minha participação na pesquisa e concordo em participar.**
**O pesquisador me informou que o projeto foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da UFSCar que funciona na Pró-Reitoria de Pós-Graduação e Pesquisa da Universidade Federal de São Carlos, localizada na Rodovia Washington Luiz, Km. 235 - Caixa Postal 676 - CEP 13.565-905 - São Carlos - SP – Brasil. Fone (16) 3351-8110. Endereço eletrônico: cephumanos@power.ufscar.br**

_____ _____
Assinatura do Participante Data

As informações referentes aos participantes foram omitidas, estando anexadas ao relatório com as respostas dos questionários devidamente respondidos, datados e assinados enviados ao Comitê de Ética.

### 6.4.3. Questionário

Para cada participante foi apresentado o termo de consentimento para que assinasse se estivesse de acordo com as diretivas declaradas. As próximas etapas constituíram da explicação e execução da "PrivPerson", permitindo que o participante a utilizasse. Após a utilização da "PrivPerson" foi aplicado um questionário, de modo que os participantes pudessem respondê-lo, baseados nas observações feitas durante a interação com a "PrivPerson".

Durante a fase de uso da "PrivPerson" o usuário pôde definir qual mecanismo avaliar primeiro, e para a avaliação dos sites puderam ser feitos testes com sites de interesse dele, já que o objetivo da pesquisa era medir a satisfação encontrada pelos usuários depois de utilizá-la. No entanto, foi ressaltado que um dos sites a serem avaliados deveria estar enquadrado na categoria do comércio eletrônico, onde as coletas de dados pessoais dos usuários são de extrema relevância, por medidas de oferta de serviços personalizados e estatísticas, o que pode em vezes invadir a privacidade.

O questionário apresentado aos usuários foi composto de 3 temas correlatos, o que no total consistiu de 17 perguntas referentes à navegação, privacidade, personalização e à utilização da "PrivPerson", sendo mostrado na Tabela 8.

**Tabela 8. Questionário apresentado para avaliação da "PrivPerson"**

| **Questionário de Avaliação** | | |
|---|---|---|
| **a) Perguntas Gerais Referentes à Navegação na Web:** | **Sim** | **Não** |
| 1 – É importante poder navegar pela web sem ser identificado? | | |
| 2 – É relevante ter garantia de que seus dados não serão coletados? | | |
| 3 – É agradável poder utilizar serviços personalizados? | | |
| 4 – Você confia no que é descrito nas Políticas de Privacidade dos sites? | | |
| **b) Sobre a Coleta de Dados Durante a Navegação pela Web:** | | |
| 1 – É importante ter controle sobre a coleta de seus dados pessoais? | | |
| 2 – É importante ter conhecimento sobre o que é feito com seus dados? | | |
| 3 – É importante saber a quem seus dados pessoais são passados? | | |
| **c) Perguntas Referentes à Utilização da Ferramenta "PrivPerson":** | | |
| 1 - Traz clareza quanto ao nível de privacidade oferecido pelos mecanismos? | | |
| 2 – Traz clareza quanto ao nível da personalização disponibilizada pelos mecanismos? | | |
| 3 – Oferece um ambiente amigável? | | |
| 4 – Apresenta informações relevantes sobre privacidade e personalização? | | |
| 5 – Apresenta informações de como utilizar a "PrivPerson" e qual seu propósito? | | |
| 6 – A utilização da ferramenta foi satisfatória? | | |
| 7 – Existem itens que deveriam ser melhorados na ferramenta? | | |
| 8 – A "PrivPerson" possibilitou que você, a partir de agora, preste mais atenção na segurança de suas informações, tomando mais cuidado durante sua navegação pela web? | | |
| 9 – A "PrivPerson" fez com que você valorize mais os serviços de personalização que são oferecidos pelos mecanismos? | | |
| 10 – A implementação da "PrivPerson" foi relevante para aumento de sua segurança? | | |
| **d) Espaço Destinado às Justificativas, caso haja necessidade:** | | |
| Obs. Se a justificativa for referente a uma das questões acima, favor indicar a letra referente ao tema e o número da pergunta.<br>_____<br>_____ | | |
| Segue abaixo as assinaturas para validade e veracidade das respostas dadas acima.<br><br>_____<br>Participante da Pesquisa<br><br>_____<br>Luanna Lopes Lobato<br>Responsável pela Pesquisa<br><br>São Carlos, ___ de _____ de 2007. | | |

As respostas ao questionário restringiram-se nas opções “Sim” ou “Não” para facilitar o preenchimento e simplificar o entendimento dos conceitos. Não foi julgado obrigatório ressaltar a justificativa pela escolha da resposta, já que as perguntas são dispostas de forma clara. No entanto, caso o participante considerasse relevante e quisesse justificá-las, foi possível.

A tabulação dos resultados para cada questão é apresentada no próximo capítulo, onde são ressaltados os resultados encontrados com o desenvolvimento de todo o trabalho.

## 6.5. Conclusões

O desenvolvimento da “PrivPerson” foi importante para validar conceitos sobre a privacidade e a personalização oferecida durante a interação dos usuários com a web. Através do uso da “PrivPerson” pode-se utilizar, de forma prática, as taxonomias definidas, verificando nos mecanismos quais das camadas são contempladas.

Sua implementação ainda foi necessária devido ao fato de poder se ter um repositório de informações sobre as ferramentas que oferecem privacidade e personalização aos usuários, facilitando dessa forma o estudo dos pesquisadores dessa área. Além de ser também um repositório de cenários de utilização e suas respectivas características e análises.

Outra vantagem apresentada pela “PrivPerson” é a sua capacidade de avaliação automatizada, principalmente em relações a sites web os quais estão sujeitos a constantes modificações, o que minimiza o tempo de busca e análise por assuntos nos sites.

A “PrivPerson” aborda temas referentes a privacidade e personalização, permitindo que os usuários estejam integrados do assunto e percebam o quão importante é ter a preocupação com sua privacidade e os benefícios oferecidos pela personalização.

Para disponibilização das informações sobre a “PrivPerson”, sobre suas funcionalidades, utilização e acesso ao programa de instalação da ferramenta, foi desenvolvido um site, o qual encontra-se no endereço: *http://www.dc.ufscar.br/privperson*.

Tal site foi desenvolvido com o objetivo de divulgar a “PrivPerson” e ainda colher resultados sobre as opiniões dos usuários após utilizá-la.

Para manter um histórico sobre as opiniões dos usuários, no site ainda é disponibilizado o questionário, apresentado anteriormente na Tabela 8, para avaliação da “PrivPerson”. As perguntas dos questionários são respondidas por meio da seleção de resposta e um campo texto para serem submetidas as justificativas, sugestões e reclamações sobre a “PrivPerson”.

Antes que o questionário seja respondido é disponibilizada para os usuários uma página contendo o termo de consentimento livre e esclarecido, também já apresentado na Tabela 7, o qual traz aos usuários informações de esclarecimento sobre o preenchimento do questionário. Ao final do termo encontra-se o botão para redirecionamento do usuário à página que disponibiliza o questionário.

No próximo capítulo é apresentado o resultado da utilização da "PrivPerson" a partir da aplicação do teste com alguns usuários selecionados, de modo a levantar mais dados concretos sobre sua utilização e relevância.

Ainda no Capítulo 7, é apresentada como resultado final uma comparação da "PrivPerson" com o Estudo de Caso, apresentado no APÊNDICE A, fazendo uma comparação da avaliação automatizada com a manual.

# 7. Resultados Finais

A mais relevante contribuição apresentada pelo trabalho realizado é a definição das taxonomias de privacidade e personalização. Essas são constituídas de camadas, as quais representam características relevantes que devem ser analisadas nos mecanismos que oferecem privacidade e personalização aos usuários.

Essas taxonomias foram impostas de modo que sejam utilizadas para medir a privacidade dos usuários e a disponibilização de serviços personalizados, otimizando as buscas e proporcionando maiores facilidades durante a navegação dos usuários na Internet.

Desenvolveu-se ainda neste trabalho um Estudo de Caso, apresentado no APÊNDICE A, no qual as taxonomias podem ser aplicadas, de modo a verificar suas relevâncias.

Para essa avaliação, as camadas foram descritas em itens que correspondem a características menos abstratas, pois apenas a utilização das camadas tornaria a análise subjetiva.

Com os resultados encontrados pelo Estudo de Caso, foi possível observar que não existia uma padronização para as Políticas de Privacidade disponibilizadas pelos sites aos usuários. Dessa forma, tornou-se possível o desenvolvimento de um padrão de Política de Privacidade, o qual aborda as características apresentadas pelas camadas das taxonomias, de modo que nos sites possa ser seguido um padrão para definição de suas políticas. Tal padronização pode ser visualizada no APÊNDICE B.

Outra grande relevância apresentada pela pesquisa é o desenvolvimento de uma ferramenta, chamada de “PrivPerson”, a qual, através do uso das taxonomias, retorna aos usuários o nível de privacidade e personalização oferecido pelos mecanismos, sendo esse nível um valor quantificado a partir dos pesos atribuídos às camadas.

Tendo os resultados encontrados para análise de sites na aplicação do Estudo de Caso, na avaliação manual, tornou-se possível fazer uma comparação com os resultados obtidos com a utilização da “Privperson”, na avaliação automatizada. Essa comparação é apresentada nas próximas seções.

Para comprovar a validade da “PrivPerson” desenvolveu-se um questionário, o qual foi aplicado a usuários de forma a avaliar sua eficiência e usabilidade. Tais respostas ao questionário, também apresentadas na próxima seção, seguiram uma metodologia centrada no usuário, de forma a direcionar questões que os usuários podem responder facilmente apenas pela percepção quando em interação com a “PrivPerson”, sendo tais questões relevantes ao domínio da pesquisa.

Ainda como resultado desta pesquisa foi desenvolvido um site[24], o qual permite que os usuários façam através da web a avaliação da "PrivPerson" respondendo ao questionário.

No site é disponibilizado para *download* o arquivo executável, e informações sobre a "PrivPerson", tais como seu objetivo e o propósito de sua implementação. Os resultados do questionário *online* são dispostos em um banco de dados, sendo o acesso restrito aos administradores, com o objetivo de obter estatísticas sobre as respostas dos usuários.

## 7.1. Aplicação do Questionário

Antes da aplicação do questionário aos participantes, procurou-se esclarecer todas as dúvidas surgidas durante a utilização da "PrivPerson". Isso foi feito para que a explicação não interferisse nas respostas dadas pelos usuários e os mesmos não se sentissem constrangidos, estando próximos a pesquisadora responsável, se achassem que deveriam registrar alguma advertência em relação a "PrivPerson".

Na Tabela 9 apresenta-se o número total de respostas para cada uma das questões respondidas pelos 20 participantes selecionados.

**Tabela 9. Respostas dos participantes ao questionário**

| Seções | Item | Sim | Não |
|---|---|---|---|
| **a) Perguntas Gerais Referentes à Navegação na Web** | 1 – É importante poder navegar pela web sem ser identificado? | 19 | 1 |
| | 2 – É relevante ter garantia de que seus dados não serão coletados? | 19 | 1 |
| | 3 – É agradável poder utilizar serviços personalizados? | 20 | 0 |
| | 4 – Você confia no que é descrito nas Políticas de Privacidade dos sites? | 5 | 15 |
| **b) Sobre a Coleta de Dados Durante a Navegação pela Web** | 1 – É importante ter controle sobre a coleta de seus dados pessoais? | 20 | 0 |
| | 2 – É importante ter conhecimento sobre o que é feito com seus dados? | 20 | 0 |
| | 3 – É importante saber a quem seus dados pessoais são passados? | 20 | 0 |
| **c) Perguntas Referentes à Utilização da Ferramenta "PrivPerson"** | 1 – Traz clareza quanto ao nível de privacidade oferecida pelos mecanismos? | 20 | 0 |
| | 2 – Traz clareza quanto ao nível de personalização disponibilizada pelos mecanismos? | 20 | 0 |
| | 3 – Oferece um ambiente amigável? | 20 | 0 |
| | 4 – Apresenta informações relevantes sobre privacidade e personalização? | 20 | 0 |
| | 5 – Apresenta informações de como utilizar a "PrivPerson" e qual seu propósito? | 20 | 0 |
| | 6 – A utilização da ferramenta foi satisfatória? | 20 | 0 |
| | 7 – Existem itens que deveriam ser melhorados na ferramenta? | 8 | 12 |

[24] http://www.dc.ufscar.br/privperson

| | | | |
|---|---|---|---|
| | 8 – A “PrivPerson” possibilitou que você, a partir de agora, preste mais atenção na segurança de suas informações, tomando mais cuidado durante sua navegação pela web? | 19 | 1 |
| | 9 – A “PrivPerson” fez com que você valorize mais os serviços de personalização que são oferecidos pelos mecanismos? | 18 | 2 |
| | 10 – A implementação da “PrivPerson” foi relevante para aumento de sua segurança? | 18 | 2 |
| **d) Espaço Destinado às Justificativas** | **Justificativas** | | |
| | | | |

Durante o preenchimento do questionário pelos participantes, pôde-se verificar clareza e entendimento dos mesmos quanto às questões disponibilizadas, tendo sido respondidas de forma rápida e segura.

Dentre os 20 participantes selecionados para avaliar a “PrivPerson”, 25% desses justificaram a escolha da opção para alguns dos itens. Essas justificativas apresentam-se como características positivas à “PrivPerson”, tendo apenas algumas dessas sido sugestões de melhoria da mesma. Tais justificativas são mostradas ao longo da descrição dos resultados.

Em relação a navegação na web, item a, pode-se observar que 19 participantes, 95%, acham relevante a navegação de forma anônima, tendo os mesmos ainda considerado que é necessário ter garantia de que seus dados pessoais não serão coletados. No entanto, todos os participantes desejam ter serviços personalizados.

Ainda verifica-se que os usuários não conhecem ao certo a necessidade da coleta de suas informações pessoais e a julga apenas como invasão de privacidade. Percebe-se que não é desejado pelos usuários que os dados sejam coletados, mas eles desejam que sejam oferecidos serviços de personalização.

Apenas 25% dos usuários confiam no que é descrito nas Políticas de Privacidade dos sites, o que implica dizer que os usuários utilizam dos sites de maneira desconfortável.

Em relação ao item b, coleta de dados durante a navegação pela web, pode-se observar que os 20 participantes consideraram importante ter controle sobre a coleta de seus dados pessoais e ainda acharam necessário saber o que é feito com os dados e a quem são passados.

Esses resultados mostram que os usuários estão preocupados com a segurança de suas informações, com a integridade e veracidade das mesmas, sendo que a possibilidade de privacidade e segurança dos dados pessoais é um assunto que já encontra-se formado nas mentes dos usuários.

Quanto ao item c, onde apresentam-se perguntas referentes à utilização da "PrivPerson", as quais são as de maior interesse neste estudo, pode-se verificar que sua utilização, desenvolvimento e eficiência mostrou-se relevante aos usuários.

Observa-se que os usuários entenderam o sentido da quantificação das camadas para saber o nível de privacidade e personalização disponibilizado pelos mecanismos, pois todos os participantes responderam que a "PrivPerson" traz clareza quanto ao nível de privacidade e personalização oferecido. Todos os participantes também relatam que a mesma oferece um ambiente amigável.

Além das respostas positivas quanto ao item que trata da coleta de dados, pode-se comprovar que os participantes estão interessados nesse assunto, pois todos responderem que a "PrivPerson" apresenta informações relevantes sobre privacidade e personalização.

Comprova-se, dessa forma, que os participantes não apenas utilizaram-na com vistas a responder o questionário, mas também navegaram por entre as funcionalidades disponibilizadas pela "PrivPerson" e leram sobre as informações referentes à privacidade e à personalização contempladas pela mesma.

Além das informações sobre os assuntos de privacidade e personalização disponibilizadas pela "PrivPerson", essa também apresenta informações que servem de guias aos usuários. Isso pode ser comprovado através das 20 respostas positivas de que a "PrivPerson" disponibiliza informações de como utilizá-la e referente ao seu propósito.

Baseados na aplicação, utilização, relevância e funcionalidade disponibilizada pela "PrivPerson", 100% dos usuários consideraram sua utilização satisfatória.

Dentre os participantes selecionados, 12 desses, 60%, consideraram que não existem itens a serem melhorados na "PrivPerson", no entanto, 40% opinam sobre possíveis melhorias, justificando como: *"Toda ferramenta pode ser melhorada."*; *"As definições a respeito de padrões e protocolos (P3P, por exemplo) poderiam conter mais informações"*. Não sendo tais justificativas um ponto fraco no desenvolvimento e eficiência da "PrivPerson", mas apenas melhorias que poderiam ser acopladas a ela, já tendo essas sugestões sido atendidas.

Todos os usuários consideraram a "PrivPerson" como uma ferramenta de auxílio à privacidade e personalização oferecida. Dentre os participantes, 95% responderam que a utilização da "PrivPerson" fez com que eles viessem a prestar mais atenção na segurança de suas informações, tomando mais cuidado durante a navegação pela web.

Dos 20 participantes selecionados, 18 desses, 90%, consideraram que a "PrivPerson" os instruiu quanto à importância da personalização, fazendo com que os mesmos valorizem os serviços personalizados que são oferecidos pelos mecanismos.

Também 90% desses consideram o desenvolvimento da “PrivPerson” importante para o auxílio à navegação na web, para o aumento de sua segurança e para o conhecimento de questões que antes não eram levadas em consideração durante a interação com os sites.

Pode-se observar que os usuários entenderam a relevância em se utilizar sites que especificam uma Política de Privacidade clara e que disponibilizam técnicas de privacidade e serviços de personalização, de modo a trazer maior segurança, conforto e simplicidade durante sua navegação na web.

A aplicação desse questionário aos usuários foi relevante, pois assim pode-se verificar de forma prática o quão importante a ferramenta desenvolvida é, sendo possível validar seu desenvolvimento, o que nos traz maiores seguranças sobre a relevância do trabalho desenvolvido.

## 7.2. Comparação da Avaliação Manual e Automatizada

A partir do Estudo de Caso, APÊNDICE A, obteve-se informações que possibilitaram que a análise feita nos sites de comércio eletrônico, inicialmente manual, pudesse ser comparada com uma análise automatizada feita a partir da “PrivPerson”.

Essa comparação, no resultado final, foi feita de modo a verificar a eficiência da “PrivPerson” para a avaliação dos sites e ainda comprovar a utilização das camadas das taxonomias de privacidade e personalização.

Tal análise é possível, pois para ambas as avaliações foram determinadas métricas análogas de análise, verificando os sites pelo modo de largura. No entanto, não são iguais devido à própria característica da análise, onde através da avaliação manual pode-se ter uma percepção de quais *links* seguir, não sendo analisados *links* que não pareçam ser relevantes.

Com a análise manual, além da demora em ter que analisar os sites, essa pode causar insatisfação aos usuários pois a busca pelo que se deseja pode tornar-se cansativa e desmotivante, principalmente quando os sites não seguem um padrão para definição de suas políticas, onde a maioria das informações encontram-se disponíveis.

Dessa forma, apresenta-se no APÊNDICE B uma padronização para as Políticas de Privacidade, de modo que essas sejam de fácil localização pelos usuários nos sites e de fácil entendimento, abordando assuntos relevantes sobre a garantia de privacidade e oferta de serviços personalizados aos usuários.

Com a análise automatizada a satisfação do usuário durante a navegação pelo site tende a ser maior, pois esse tem a possibilidade de conhecer as práticas dos sites com maior

precisão. Com essa análise, diminui-se o tempo de busca pelo conhecimento de algum tipo de informação disponibilizado ou não pelos sites e ainda, torna-se possível a análise de todas as URL's referenciadas pelos sites, o que é inviável na análise manual para sites que contêm referência a muitos *links*.

Assim como as informações pessoais dos participantes da pesquisa foram mantidas em sigilo, informações de identificação dos sites também não são disponibilizadas, de modo que esta pesquisa não represente marketing a nenhum deles, e sim, seja focada na comparação da avaliação manual e automatizada.

Nas tabelas a seguir são mostrados os resultados da análise manual e da automatizada, sendo apresentando a porcentagem de sites, dentre os 33 sites selecionados para a amostra, que contemplam as camadas das taxonomias de privacidade e personalização.

Na Tabela 10são apresentados os resultados da análise manual para a taxonomia de privacidade, a qual consistiu em uma inspeção manual nos sites de comércio eletrônico em busca das camadas que esses contemplam, sendo os detalhes dessa disponibilizados no APÊNDICE A. Também são apresentados os resultados da análise automatizada para as camadas da taxonomia de privacidade, feita com o uso da "PrivPerson". Nessa análise a "PrivPerson" foi responsável por verificar nas páginas dos sites se esses utilizavam as camadas da taxonomia de privacidade, de forma automatizada, retornando ao usuário o nível de privacidade oferecido pelo site em análise.

**Tabela 10. Resultado das análises para a taxonomia de privacidade**

| Número da Camada | Camada Analisada | Resultado da Análise Manual | Resultado da Análise Manual |
|---|---|---|---|
| 1 | Leis de Proteção de Privacidade | 0% | 0% |
| 2 | Políticas de Privacidade | 75,75% | 81,81% |
| 3 | Certificação de Privacidade | 48,48% | 75,75% |
| 4 | Notificação | 57,57% | 84,84% |
| 5 | Controle | 87,87% | 81,81% |
| 6 | Mecanismos para Proteção de Privacidade | 90,90% | 81,81% |

Pode-se observar que, tanto na avaliação manual quanto na automatizada a Camada 1 - Leis de Proteção de Privacidade não foi contemplada por nenhum dos sites da amostra selecionada. Isso se deu devido à amostra selecionada ser de um conjunto de sites de comércio eletrônico brasileiro e, como já mencionado, no Brasil não existem leis específicas à proteção de privacidade dos usuários na web.

A Camada 2 - Política de Privacidade teve um maior percentual de encontro na avaliação automatizada, pois nessa são analisados todos os *links* referenciados pelo site, sendo

o *link* da Política de Privacidade reconhecido durante essa análise. Já na avaliação manual, a busca por tais políticas torna-se cansativa devido aos sites não seguirem uma padronização para sua definição e disponibilização, o que pode dificultar a localização dessas pelos usuários.

O uso da Camada 3 - Certificação de Privacidade apresentou uma significativa diferença na avaliação manual com a automatizada, pois na avaliação manual foi encontrado uma porcentagem menor de sites que utilizam certificação. Isso se dá pois alguns sites não disponibilizam informações aos usuários de forma clara sobre o uso de entidades certificadoras.

Através da avaliação manual, pode-se encontrar um número maior de sites que apresentam a Camada 5 - Controle. Isso ocorre devido ao fato de que esse tipo de serviço pode ser descrito nos sites das mais diversas formas, podendo ser reconhecidas pela percepção humana, já que não existe uma padronização para desenvolvimento dos sites. Dessa forma, a "PrivPerson" pode não reconhecer todas as formas como essa funcionalidade é disponibilizada.

A seguir são apresentadas as tabelas da avaliação manual e automatizada para a ocorrência da utilização das camadas da taxonomia de personalização. Na Tabela 11 é mostrada a porcentagem de sites que apresentaram cada dessas camadas na avaliação manual e na automatizada.

**Tabela 11. Resultado das análises para a taxonomia de personalização**

| Número da Camada | Camada Analisada | Resultado da Análise Manual | Resultado da Análise Automatizada |
|---|---|---|---|
| 1 | Controle de Filtros | 78,78% | 81,81% |
| 2 | Coleta de Dados | 96,96% | 84,84% |
| 3 | Transferência de Dados | 84,84% | 78,78% |
| 4 | Utilização de Intermediário | 66,66% | 60,60% |
| 5 | Armazenamento em Servidor | 48,48% | 69,69% |
| 6 | Históricos | 51,51% | 78,78% |

Uma das vantagens da avaliação automatizada é sua dinamicidade, o custo de se fazer novas medições ao longo do tempo é fixo, sendo o custo do desenvolvimento inicial diluído nessas novas medições. Por outro lado, na avaliação manual, se houver a necessidade de se fazer uma avaliação mais atual, todo o intenso trabalho deve ser refeito.

Tanto os resultados da avaliação manual quanto da automatizada são válidos em um espaço de tempo consideravelmente pequeno, já que os sites estão em constantes modificações. No entanto, julga-se mais relevante fazer a avaliação automatizada, onde o

trabalho é passado todo ao software de análise, visto que na avaliação manual, feita pelo usuário, a análise é demorada, cansativa e propensa a erros.

Deve-se considerar, no entanto, que como já apresentado nas tabelas acima, algumas das características apresentadas pelas camadas das taxonomias de privacidade e personalização podem não ser encontradas com a mesma porcentagem que apresentada na avaliação manual. Isso se dá, pois existem muitas possibilidades de como as características que representam as camadas estão sendo dispostas pelos sites, e podem não ter sido fornecidos à "PrivPerson" todas as possibilidades para encontrar essas características.

No entanto, o não reconhecimento dessas características através do uso da "PrivPerson" não ocorre com freqüência, já que o *crawler* da mesma tem opções de palavras e utiliza técnicas computacionais para o reconhecimento dessas, ocasionando na identificação das camadas das taxonomias.

Mesmo que na avaliação manual para algumas camadas tenha sido encontrada uma porcentagem maior de ocorrência dessas nos sites, tais resultados não inviabilizam a utilização da "PrivPerson" para a análise automatizada, já que essa pode ser abastecida com informações pertinentes que auxiliem na avaliação dos sites e identificação das camadas.

Outras características referentes às camadas que podem não ter sido encontradas durante a avaliação automatizada e, no entanto foram encontradas na análise manual, ocorre devido a funcionalidades que os sites empregam como requisitos para acesso, tendo os usuários que se cadastrarem para acessar algumas informações, sob a forma de login.

Através da avaliação manual podem-se observar algumas características não encontradas na automatizada, pois foi possível efetivar o cadastro no sites e constatar a utilização das camadas, apesar da demora na efetivação do cadastro e na busca por tais características.

Apesar de automatizar o acesso e a busca às camadas das taxonomias, a "PrivPerson" não tem a capacidade de realizar cadastros, pois esses são dispostos de diferentes maneiras pelos sites. Essa é uma limitação apresentada pela "PrivPerson", o que pode em vezes prejudicar a avaliação dos sites quando esses utilizam da obrigação de cadastros para ter acesso às informações disponibilizadas.

Alguns sites ainda utilizam o bloqueio de cadastros automáticos por *crawlers* na web, de modo a evitar que usuários falsos, ou seja, clientes inexistentes acessem seus sites, sendo necessário fazer autenticação de usuário por meio de login e senha, para acessar algumas páginas. Outros também disponibilizam ao usuário um código que deve ser informado durante

a efetivação do cadastro, de modo a trazer maior segurança de que é um humano e não robôs da web que estão se cadastrando no site.

No entanto, pode-se observar nas tabelas, através da porcentagem atribuída às camadas, que essas foram identificadas nos sites analisados durante a avaliação automatizada. Isso ocorre pois a maioria dos sites disponibilizam informações aos usuários antes da efetivação de cadastros, com o objetivo de atrair os usuários aos sites.

Tais informações são na maioria das vezes disponibilizadas nas Políticas de Privacidade dos sites. Devido a isso vê-se a importância em desenvolvê-las de forma claras e de fácil acesso aos usuários.

As figuras a seguir mostram, na forma de gráficos, o nível de privacidade e personalização tido como resultado pela utilização das camadas pelos 33 sites para os dois tipos de análise. Na Figura 35 é retratada a análise manual, onde os sites foram avaliados de acordo com a percepção humana.

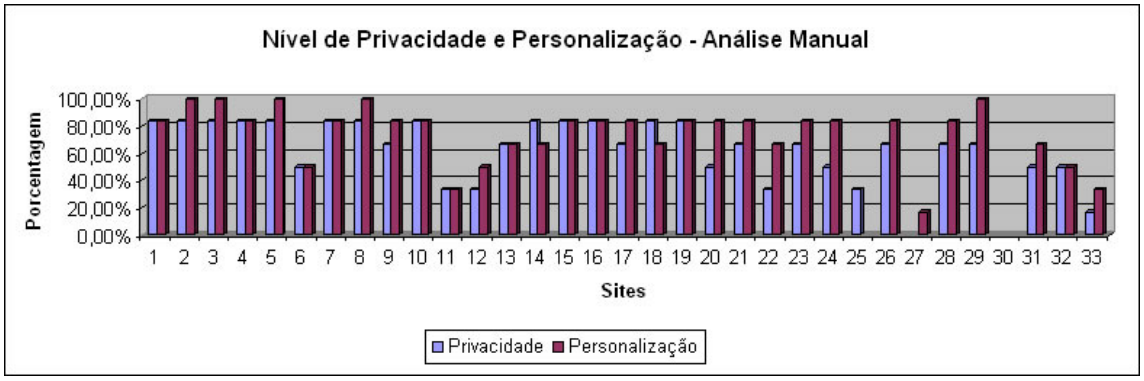


**Figura 35. Nível de privacidade e personalização – análise manual**

Tal demonstração foi feita de modo a melhor visualizar a diferença encontrada em cada uma das análises e demonstrar a situação dos sites de comércio eletrônico brasileiro.

Na Figura 36 é apresentado o gráfico para a análise automatizada, a qual a avaliação nos sites foi feita com o uso da "PrivPerson".

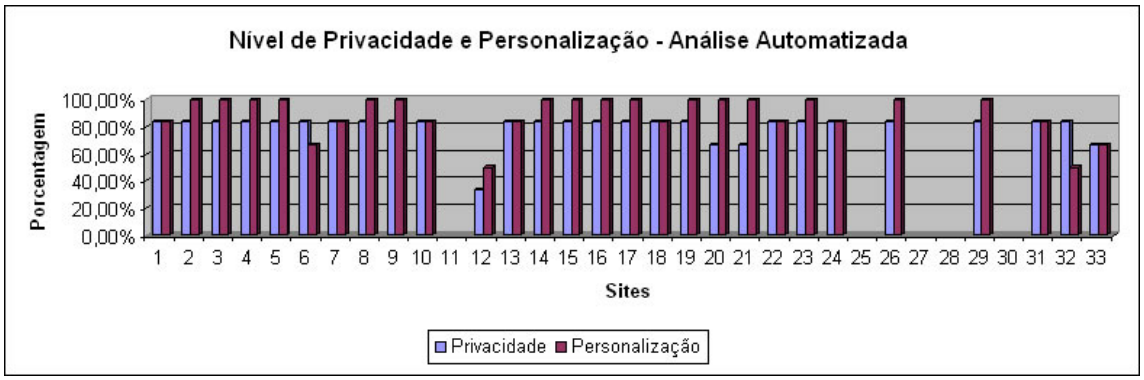


**Figura 36. Nível de privacidade e personalização – análise automatizada**

Observa-se que, de acordo com as taxonomias de privacidade e personalização, os sites em geral têm apresentado de forma significativa as camadas, o que foi a base para a determinação dos pesos. No entanto, a utilização de algumas camadas ainda não encontra-se tão satisfatória, principalmente na análise manual.

Percebe-se com a comparação dos gráficos que na análise automatizada uma quantidade maior de sites não pôde ser analisada. Isso se deu devido ao fato de que esses são desenvolvidos utilizando *javascrip*, o que impossibilitou a análise. Também, como já mencionado, para alguns sites o nível de privacidade e personalização apresentado na análise automatizada foi menor que na manual devido ao fato de alguns sites necessitarem de cadastro para acesso às suas páginas.

No entanto, a "PrivPerson" encontrou alguns itens que não haviam sido observados durante a análise manual, apresentando dessa forma, a análise automatizada uma grande otimização na busca e na satisfação da pesquisa.

Comprova-se com mais essa representação a viabilidade e importância da "PrivPerson" e de todo o trabalho desenvolvido, já que estão todos os resultados ligados intrinsecamente uns aos outros para validade dos demais, tendo sido utilizados em conjunto para completar o trabalho como um todo.

No entanto, esses resultados foram disponibilizados principalmente de modo a ressaltar a importância na utilização das camadas das taxonomias de privacidade e de personalização, por essas serem a base da contribuição apresentada pelo trabalho, pois a partir delas pode-se desenvolver todo o trabalho.

## 7.3. Conclusões e Trabalhos Futuros

O trabalho desenvolvido contribui de modo significativo para a área de privacidade e personalização, trazendo uma fonte de pesquisa, de experimentos realizados e resultados obtidos, onde os usuários e pesquisadores poderão basear-se para o desenvolvimento de outras propostas similares.

Neste trabalho, importantes decisões de construção de projetos e implementações foram tomadas e colocadas em prática, tendo a "PrivPerson" como grande aliada ao tema em questão.

As taxonomias apresentadas apresentaram uma relevante utilização para garantia da privacidade e oferta de serviços personalizados aos usuários, e com a utilização dessas pela "PrivPerson", pode-se medir o nível de privacidade e personalização oferecido pelos mecanismos analisados.

Comprova-se através do estudo feito com os usuários, o quão importante foi o desenvolvimento da "PrivPerson" e, o quanto de benefícios ela oferece, mesmo que em uma versão inicial, a qual, no entanto, já contempla e supre o trabalho de mestrado proposto.

Através do uso da "PrivPerson" pode-se aplicar de forma prática a importância da privacidade e personalização durante a interação dos usuários com a web, quantificando o nível apresentado.

A "PrivPerson" é *open source*, estando disponível para testes e maiores contribuições de funcionalidades no endereço *http://www.dc.ufscar.br/privperson*, com o termo de consentimento livre e esclarecido e o questionário a ser respondido após sua utilização, com vistas a medir seu funcionamento e importância.

No entanto, deve ser observado que para sua utilização deve-se seguir a licença GPL (*General Public License*) a ela atribuída. É permitida sua utilização, ou modificação se necessário, sem que os autores percam o direito autoral e de forma que a ferramenta continue estando aberta à comunidade de pesquisa e desenvolvimento.

Neste trabalho foram utilizadas *threads* para análise dos sites, no entanto, podem ser utilizadas diferentes técnicas para facilitar e otimizar a avaliação desses. Essas novas funcionalidades, acopladas à "PrivPerson" irão otimizar os resultados tanto em integridade das informações a serem analisadas quanto em velocidade de acessos e respostas aos usuários.

Como trabalhos futuros pode-se citar um melhoramento nas funcionalidades atribuídas à "PrivPerson", de modo que as análises dos mecanismos sejam melhoradas. Algumas sugestões de trabalho futuros são:

- Permitir que a "PrivPerson" utilize heurísticas próprias para se cadastrar em sites e assim analisar páginas que demandem um acesso autenticado;
- Implementar um compilador javascript para que a "PrivPerson" seja capaz de analisar sites em que os *links* estejam implementados em javascript;
- Permitir a atualização dos arquivos XML através da web;
- A partir das páginas que o *crawler* faz *download*, aplicar técnicas de recuperação de informação (modelo booleano, vetorial ou probabilístico) para otimizar as buscas por palavras relevantes para a consulta da utilização das camadas das taxonomias;
- Possibilitar que os usuários guardem as URL's já analisadas em uma seção Favoritos, possibilitando o acesso fácil aos sites que retornarem um nível de privacidade e personalização satisfatório;

- Propõe-se ainda um melhoramento na interface da “PrivPerson” com vistas a cada vez torná-la mais amigável ao usuário, facilitando sua utilização;
- Ainda é sugerido como trabalhos futuros o desenvolvimento da “PrivPerson”, da parte de análise de sites, como um *plug-in* para os navegadores ou um *tray*, acoplados à barra de atividades no ambiente de trabalho dos sistemas operacionais.

Tal pesquisa traz como benefício a colaboração para o meio de acadêmico, contribuindo para um avanço tecnológico, onde toda a comunidade de usuários que utilizam a web, assim como pesquisadores da área e área correlatas, poderão se beneficiar com a pesquisa.

Durante o desenvolvimento do trabalho de mestrado e posterior escrita da dissertação, foi possível obter a aprovação de artigos referentes às pesquisas desenvolvidas, validando dessa forma o estudo apresentado.

# Referências Bibliográficas


AGUIAR, A. **Proteção na rede**. 2006. Revista Consultor Jurídico. 9 de julho de 2006. Disponível em: <http://conjur.estadao.com.br/static/text/46139,1>. Acesso em: 02 março 2007.

ANNIE, I. *et al.* **The lack of clarity in financial privacy policies and the need for standardization**. IEEE Security & Privacy, 2(2), pp. 36-45 2004.

BEIZER, B. **Software testing techniques**. 2ª Ed. . New York: Van Nostrand Reinhold Company. 1990.

BOOCH, G.; RUMBAUGH, A. e JACOBSON, I. **UML: Guia do Usuário**. v 1. São Paulo: Editora Campus. 2005. 474 p.

BORCHERS, J. **A Pattern Approach to Interaction Design**. In: Conference on Designing interactive systems: processes, practices, methods, and techniques.New York, United States: 2000. Disponível em: <http://portal.acm.org/citation.cfm?id=558433&coll=Portal&dl=GUIDE&CFID=3720139&CFTOKEN=24769028#>. Acesso em: 16 out. 2006.

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

BUGLIESI, M. e GIUNTI, M. **Secure implementations of typed channel abstractions**. In: Proceedings of the 34th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. Nice, France: 2007. Pág. 251-262.Disponível em: <http://doi.acm.org/10.1145/1190216.1190253>. Acesso em: 03 março 2007.

CHAUM, D. **Untraceable electronic mail, return addresses and digital pseudonyms**. In: Commun. ACM 24, 2 (Feb. 1981). 1981. Pág. 84-90.Disponível em: <http://doi.acm.org/10.1145/358549.358563>. Acesso em: 29 nov. 2005.

CHELLAPPA, R. K. e SIN, R. **Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma**. 2005. Department of Information and Operations Management, Marshall Scholl of Business, University of Suthern California. LA. Disponível em: <http://asura.usc.edu/~ram/rcf-papers/per-priv-itm.pdf>. Acesso em: 25 jan. 2006.

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

CRANOR, L. *et al.* **The Platform for Privacy Preferences 1.0 (P3P1.0) Specification**. 2002. World Wide Web Consortium Recommendation. Disponível em: <http://www.w3.org/TR/P3P/>. Acesso em: 8 set. 2005.

CRANOR, L. F.; BYERS, S. e KORMANN, D. **An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites**. 2003. 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.

CRETELLA, J. J. **Comentários à Constituição Brasileira de 1988**. Rio de Janeiro, RJ, v 1, 1997.

CROCKER, D. H. O., P. **Augmented BNF for Syntax Specifications: ABNF**. Internet proposed standard RFC4234. 2005. Disponível em: <http://dret.net/rfc-index/reference/RFC4234>. Acesso em: 26 fev. 2006.

DOYLE, B.; SHEVLIN, R. e WATSON, T. **What Satisfies Financial Services Consumers**. 2004. Disponível em: <http://www.edwardjones.com/cgi/getData.cgi?file=/pdf/new_condensed_forrester.pdf>. Acesso em: 21 jan. 2006.

E-BIT. W**eb Shoppers - Raio-X do comércio eletrônico em 2005**. 2006. Disponível em: <http://www.camara-e.net/_upload/WebShoppers13.pdf>. Acesso em: 28 jun. 2006.

ELGESEM, D. **Privacy, respect for persons, and risk**. Charles Ess, editor, Philosophical perspectives on computer-mediated communication. State University of New York Press, chaper 3, pages 45-66. 1996.

ESMIN, A. A. **Modelando com UML – Unified Modeling Language**. INFOCOMP Journal of Computer Science v.1, 1807-4545 November, 1999, p.48-58. 1999.

FAYYAD, U.; PIATETSKI-SHAPIRO, G. e SMYTH, P. **The KDD process for extracting useful knowledge from volumes of data**. In: Commun. ACM 39, 11 (Nov. 1996), 27-34. 1996.Disponível em: <http://doi.acm.org/10.1145/240455.240464>. Acesso em: 02 nov. 2005.

FERREIRA, A. B. D. H. **Dicionário Aurélio**. vol.1: 2 Ed. Rio de Janeiro: Editora Nova Fronteira. 1989. 536 p.

FIDEL, R. Q**uality Methods in Information Retrieval Research**. 1993. Disponível em: <http://www.ischool.washington.edu/fidelr/RayaPubs/QualitativeMethodsinInformationRetrievalResearch.pdf>. Acesso em: 12 jun. 2006.

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

FRIEDMAN, B.; KHAN JR., P. H. e HOWE, D. C. **Trust online**. In: Commun ACM 43, 12 (Dec. 2000), 34-40. 2000. Disponível em: <http://doi.acm.org/10.1145/355112.355120>. Acesso em: 08 out. 2005.

GABEER, E. *et al.* **LPWA Lucent personalized web assistant**. 1998. Bell Laboratories, Information Sciences Research Center, Lucent Technologies. Disponível em: <http://www.bell-labs.com/projects/lpwa>. Acesso em: 25 jan 2006.

GAERTNER, A. e SILVA, H. P. **Privacidade da Informação na Internet: Ausência de Normalização**. (Dissertação - Mestrado em Ciência da Informação). Instituto de Ciência da Informação, Universidade Federal da Bahia, Ba, 2006. p 15.

GARFINKEL, S. **Web Security, Privacy & Commerce**. 2nd edition. O'Reilly, jan. 2002.

GERBER, L. D. e BECKER, K. **Contributions of Pattern Languages to Framework-Based Development in Layered Architectures**. In: XXVI Conferencia Latino Americana de Informática - CLEI Mexico: 2000. Disponível em: <http://www.inf.ufrgs.br/~nina/ListPub.html>. Acesso em: 12 nov. 2006.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4ª Ed. São Paulo: Atlas. 2002. 171 p.

GOLDBERG, I.; WAGNER, D. e BREWER, E. **Privacy-enhancing technologies for the Internet**. In: Proc. of IEEE Spring CompCon. 1997. Disponível em: <http://citeseer.nj.nec.com/54687.html>. Acesso em: 03 dez. 2005.

HALL, C. S. e LINDZEY, G. **Theories of Personality**. John Wiley & Sons 3 ed. 1978.

HAN, J. e KAMBER, M. **Data Mining: concepts and techiniques**. 1. New York: Morgan Kaufmann. 2001.

HARPER, C. **The Daily Me**. American Journalism Review. 1997.

HARTLEY, J. F. **Case studies in organizational research**. In: Qualitative methods in organizational research: a practical guide. London, v, 1994, p.253 p. 208-229.

HOLZINGER, A. **Usability Engineering Methods for Software Developers**. Communications of the ACM Vol. 48 No 1, 2005. 71-74 p.

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

ISHITANI, L.; ALMEIDA, V. e MEIRA JR., W. **MASKS: Bringing Anonymity and Personalization Together**. In: IEEE Security & Privacy. 2003. Disponível em: <http://ieeexplore.ieee.org/xpl/abs_free.jsp?arNumber=1203218>. Acesso em: 24 out. 2004.

JEVEAUX, P. C. M. **Introdução ao JavaBean**. 2004. Disponível em: <http://www.portalwebmobile.com.br/java/02.asp>. Acesso em: 13/12/06.

JUNG, C. G. **Desenvolvimento da Personalidade**. vol.XVIII: 8 ed. Petrópolis. 1998. 233 p.

KOBSA, A. **Tailoring privacy to user's needs**. vol.2109: Lecture Notes in Computer Science. 2001. 303 p.

_____. **Personalized hypermedia and international privacy**. In: Commun. ACM, 45, 5 (May. 2002). 2002. Pág. 64-67. Disponível em: <http://doi.acm.org/10.1145/506218.506249>. Acesso em: 24 nov. 2005.

KOCH, M. **User Representation in eCommerce and Collaboration Applications**. 2003. Department of Informatics, Technische Universitaet Muenchen, Germany. Disponível em: <http://www11.informatik.tu-muenchen.de/publications/pdf/Koch2003a.pdf>. Acesso em: 29 jan. 2006.

KRISTOL, D. e MONTULLI, L. **HTTP State Management Mechanism**. Bell Laboratories, Lucent Technologies. Epinions.com. RFC 2965, October 2000. 2000.

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

LAKATOS, E. M. e MARCONI, M. D. A. **Metodologia do Trabalho Científico**. São Paulo: Atlas. 2001.

LOBATO, L. L.; BITTAR, T. J. e ZORZO, S. D. **Abordagem para definição de taxonomia de Privacidade e Personalização para design de interação e gestão do conhecimento em comunidades de CSCL para Licenciatura em Computação**. In: Simpósio Brasileiro de Informática na Educação, 09 Nov. 2006. Brasília, DF: 2006. Pág. 10. Disponível em: <http://www.catolicavirtual.br/sbie2006/>. Acesso em: 15 jan. 2007.

LOBATO, L. L. e ZORZO, S. D. **Avaliação dos Mecanismos de Privacidade e Personalização na Web**. In: XXXII Conferencia Latinoamericana de Informática, CLEI, Agosto 2006. Santiago, Chile: 2006. Disponível em: <http://www.clei2006.org> Acesso em : 12 set. 2006.

_____. **Abordagem para o desenvolvimento de Padrões para Políticas de Privacidade**. DC – UFSCar (Universidade Federal de São Carlos). São Carlos, SP, p.35. 2007a. (01/2007).

_____. **Estudo de caso da avaliação por inspeção em sites de comércio eletrônico**. DC - UFSCar (Universidade Federal de São Carlos). São Carlos, SP, p.94. 2007b. (02/2007).

_____. **Padrões para apoio ao desenvolvimento de Políticas de Privacidade**. In: SugarLoafPlop'2007, Maio 2007. Porto de Galinhas, PE: 2007. 2007c. Disponível em <http://sugarloafplop.dsc.upe.br/>.

_____. **Satisfação com o Uso da Ferramenta PrivPerson para Avaliação dos Mecanismos de Privacidade e Personalização**. Comitê de Ética - UFSCar (Universidade Federal de São Carlos). São Carlos, SP, p.102. 2007d. (Comitê de Ética: 047/2007).

LUCENA, N. C. **Função social da privacidade**. 2002. Módulo Security. Disponível em: <http://www.modulo.com.br/pdf/funcaosocialpriv.pdf>. Acesso em: 19 out. 2005.

MALHOTRA, K. N. **Pesquisa de marketing: uma orientação aplicada**. vol.1: 3ª Ed. Porto Alegre: Bookman. 2001. 249 p.

MARTIN, D. **Detecting Web Bugs with Bugnosis: Privacy Advocacy through Education**. 2003. Disponível em: <http://www.bugnosis.org/faq.html>. Acesso em: 29 nov. 2005.

MATTAR, F. N. **Pesquisa de Marketing**. vol.1: 3ª Ed. São Paulo: Atlas. 2001. 275 p.

MAYER, A. H**ow to Make Personalized Web Browsing Simple, Secure, and Anonymous**. 1997. Bell Laboratories, Information Sciences Research Center, Lucent Technologies. Disponível em: <http://www.bell-labs.com/project/lpwa/papers.html>. Acesso em: 26 jan. 2006.

MCDANIEL, D. C. e GATES, R. P**esquisa de Marketing**. São Paulo: Thomson. 2003. 562 p.

MEIRA JR, W.; MURTA, C. e RESENDE, R. **Comércio Eletrônico na WWW**. XII Escola de Computação. São Paulo, SP 2000.

MESZAROS, G. e DOBLE, J. **MetaPatterns: A Pattern Language for Writing Patterns**. In: Conference on Pattern Languages of Programming PLoP. 1996. Disponível em: <http://www.hillside.net/patterns/writing/patternwritingpaper.htm>. Acesso em: 03 nov. 2006.

MILLAR, M. **Protecting privacy: Evaluating recent solutions proposed for and by the private sector**. 1995. Government Information in Canada. Disponível em: <http://www.usask.ca/library/gic/v2n1/millar/millar.html>. Acesso em: 05 nov. 2005.

MONTGOMERY, A. L. **Using Clickstream Data to Predict WWW Usage**. WebShop, University of Maryland, v 1, 2003

MOORES, T. T. **Do consumers understand the role of privacy seals in e-commerce?** In: Commun. ACM 48, 3 (Mar. 2005), 86-91. 2005. Disponível em: <http://doi.acm.org/10.1145/1047671.1047674>. Acesso em: 09 set. 2005.

MOORES, T. T. e DHILLON, G. **Do privacy seals in e-commerce really work?** In: Commun. ACM 46, 12 (Dec. 2003), 265-271. 2003. Disponível em: <http://doi.acm.org/10.1145/953460.953510>. Acesso em: 04 set. 2005.

MOTA, A. M. e GOMES, P. **Web Information Gathering - Armazenamento e Pesquisa da Informação**. Curso de Especialização em Ciências Documentais - Departamento de Letras, Universidade da Beira Interior, 2004. 11 p.

NIELSEN, J. **Risks of Quantitative Studies. Alertbox publicado em Useit.com**. 2004. Disponível em: <http://www.useit.com/alertbox/20040301.html>. Acesso em: 26 ago. 2006.

OLIVEIRA, I. R. D.; BALBY, L. e GIRARDI, R. **Padrões baseados em Agentes para a Modelagem de Usuários e Adaptação de Sistemas.** In: SugarLoafPLoP - Conferência Latino-Americana em Linguagens de Padrão para Programação.Fortaleza, Ceará: 2004. Disponível em: <sugarloafplop2004.ufc.br/acceptedPapers/ww/WW_18.pdf> Acesso em: 19 fev. 2007.

OMG. **Unified Modeling Language Specification Version 2.0.** 2006. Disponível em: <http://www.omg.org/technology/documents/formal/uml.htm>. Acesso em: 13 mar. 2006.

PAESANI, L. M. Direito e Internet: Liberdade de Informação, Privacidade e Responsabilidade Civil. 2ª Edição. 2003.

PITOFSKY, R. **Privacy online: Fair information practices in the electronic marketplace**. 2000. Federal Trade Commission. Disponível em: <http://www.ftc.gov/reports/privacy2000/privacy2000.pdf>. Acesso em: 13 nov. 2005.

PRESSMAN, S. R. **Software engineering - a practitioner's approach**. 5ª Ed.: New York: McGraw-Hill. 2001.

REGGIANI, L. **As cifras do e-Commerce**. In: Info Exame: 46 - 56 p. Num. 245. Agosto, 2006. Ano 21.

REITER, K. M. e RUBIN, D. **Crowds: anonymity for Web transactions**. In: Commun. ACM 42, 2 (Feb. 1999). 1999. Pág. 32-48. Disponível em: <http://doi.acm.org/10.1145/293411.293778>. Acesso em: 23 set. 2005.

ROCHA, B. G. **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, 2002 New York, NY: 2002. Disponível em: <http://doi.acm.org/10.1145/644527.644535>. Acesso em: 12 out. 2005.

ROMERO, C.; VENTURA, S. e BRA, P. D. **Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors**. 2005. User Modeling and User-Adapted Interaction 14, 5, 425-464. Disponível em: <http://dx.doi.org/10.1007/s11257-004-7961-2>. Acesso em: 26 out. 2005.

ROSA, E. H. P. e FANHANI, E. E. **Algumas reflexões sobre o Comércio Eletrônico: vantagens da comercialização pela Internet**. In: Revista de Administração Nobel. 03: 71-76 p. Janeiro 2004.

SHUBINA, A. M. e SMITH, S. W. **Using caching for browsing anonymity**. 2003. 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.

SILVA, A. C. D. *et al.* **Integrando visões de IHC e de ES por padróes no desenvolvimento por prototipação descartável**. 2005. ACM. Disponível em: <http://doi.acm.org/10.1145/1111360.1111383>. Acesso em: 05 out. 2006.

SILVA, J. **A. Software e privacidade: uma defesa do código-fonte aberto na preservação do direito constitucional à vida privada**. 2002. Jus Navigandi. Disponível em: <http://jus2.uol.com.br/doutrina/texto.asp?id=2931>. Acesso em: 19 set. 2005.

SPIEKERMANN, S.; GROSSKLAGS, J. e BERENDT, B. **E-privacy in 2nd Generation E-Commerce: privacy preferences versus actual behavior**. In: Proceedings of the 3rd ACM Conference on Electronic Commerce (Tampa, Florida, USA, October 14 - 17, 2001). EC '01.

ACM Press.New York, NY: 2001. Pág. 38-47. Disponível em: <http://doi.acm.org/10.1145/501158.501163>. Acesso em: 24 set. 2005.

SYVERSON, P. F.; GOLDSCHLAG, D. M. e REED, M. G. **A Formalization of Anonymity and Onion Routing**. In: IEEE Computer Society.Washigton, DC, USA: 1997. Disponível em: <http://portal.acm.org/citation.cfm?coll=GUIDE&dl=GUIDE&id=884368>. Acesso em: 05 dez. 2005.

TRAUTH, E. M. e O'CONNOR, B. **A study of the interaction between information, technology and society: an illustration of combined qualitative research methods**. In: Information Systems Research: Contemporary Approaches & Emergent Traditions. Amsterdam, v, 1991, p.131-144.

TUROW, J. **Americans and Online Privacy: The System is Broken**. 2003. Disponível em: <http://www.appcpenn.org/04_info_society/2003_online_privacy_version_09.pdf>. Acesso em: 20 jan. 2006.

VOLOKH, E. **Personalization and Privacy**. In: Commun. ACM 43, 8 (Aug. 2000). 2000. Pág. 84-88. Disponível em: <http://doi.acm.org/10.1145/345124.345155>. Acesso em: 06 out. 2005.

WANG, H.; LEE, M. K. O. e WANG, C. **Consumer privacy concerns about Internet marketing**. In: Commun. ACM 41, 3 (Mar. 1998), 63-70. 1998. Disponível em: <http://doi.acm.org/10.1145/272287.272299>. Acesso em: 03 dez. 2005.

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

YIN, R. K. **Estudo de Caso: Planejamento e método**. 3ª Ed. Porto Alegre: Bookman. 2005.

# APÊNDICE A - Estudo de Caso - Inspeção de Sites

Apresenta-se neste apêndice um Estudo de Caso em sites de comércio eletrônico, que tem como objetivo a análise desses através da inspeção manual na busca por ocorrências de características que os sites devem apresentar para garantir a privacidade e oferecer serviços de personalização aos usuários.

Essas características representam as camadas das taxonomias de privacidade e personalização, no entanto foram dispostas na forma de itens os quais são menos abstratos, sendo dessa forma possível identificá-los nos sites através da inspeção manual.

Os resultados deste Estudo de Caso serviram como base para: i) verificar a utilização das camadas da taxonomia de privacidade e personalização através da inspeção dos sites, sendo esse o objetivo principal desse apêndice; ii) comparar a avaliação manual com a automatizada, feita com o uso da "PrivPerson"; e conseqüentemente, iii) validar a eficiência na utilização da "PrivPerson", já que essa otimiza a avaliação dos mecanismos.

De acordo com Yin (2005) são empregados Estudos de Caso quando se quer investigar e compreender em profundidade fenômenos sociais, através de abordagens empíricas e holísticas de problemas contemporâneos.

Além disso, Estudos de Caso são especialmente pertinentes quando as perguntas de pesquisa buscam a compreensão de como se desenvolvem determinados processos, suas causas e motivações (como e por quê) (YIN, 2005).

Seu objetivo é compreender o evento em estudo e ao mesmo tempo desenvolver teorias mais genéricas a respeito dos aspectos característicos do fenômeno observado (FIDEL, 1993). A abordagem de Estudo de Caso não é um método propriamente dito, mas uma estratégia de pesquisa.

Segundo Hartley (1994), o Estudo de Caso consiste em uma investigação detalhada de uma ou mais organizações, ou grupos dentro de uma organização, com vistas a prover uma análise do contexto e dos processos envolvidos no fenômeno em estudo. O fenômeno não está isolado de seu contexto, já que o interesse do pesquisador é justamente essa relação entre o fenômeno e seu contexto.

Para Trauth e O'Connor (1991) uma das principais características do Estudo de Caso é que a pesquisa é dirigida aos estágios de exploração, classificação e desenvolvimento de hipóteses do processo de construção do conhecimento. Também relatam que, normalmente, uma ou mais entidades (pessoa, grupo, organização) são examinadas.

Nielsen (2004) defende o uso de estudos qualitativos com a justificativa de que esses são menos frágeis e com menor possibilidade de falhar em pontos fracos da metodologia empregada.

Nesse contexto, entende-se que o método de pesquisa que será utilizado neste trabalho é o Estudo de Caso, utilizando-se métodos qualitativos embasados em observação, marcação e tabulação de dados.

Dessa forma, o delineamento do Estudo de Caso permite contemplar o cenário de desenvolvimento do comércio eletrônico através do uso das taxonomias propostas e por evidências, com a observação de aspectos relevantes, sobre quais questões são importantes para a avaliação da privacidade e personalização oferecida pelos sites.

Os resultados são documentados e analisados visando constatar a utilização das camadas das taxonomias de privacidade e personalização, apresentadas no Capítulo 5. Os resultados ainda servem de comparação com os obtidos por uma análise automatizada, realizada através da ferramenta "PrivPerson" desenvolvida neste trabalho, os quais foram descritos no Capítulo 6.

## A.1. Levantamento de Dados

A questão da privacidade é uma das mais preocupantes no comércio eletrônico. Os sites, especificamente os de comércio eletrônico que são os retratados neste trabalho, diariamente coletam algum tipo de informação de seus usuários, desde detalhes de navegação até informações pessoais.

Alguns desses sites não têm interesse em coletar tais informações, mas o fazem em virtude da própria operação que lhes foi solicitada, utilizando os dados apenas para a efetivação do negócio. Muitos, contudo, vivem da exploração comercial dessas informações coletadas, formando grandes e valiosos bancos de dados e de consumo (ROSA e FANHANI, 2004).

Do ponto de vista jurídico, tanto a coleta não autorizada quanto a manipulação desses dados de forma indevida, não permitindo que o usuário tenha condições confortáveis para manipulação dessas informações, podem acarretar responsabilidades e danos para empresas que exploram o comércio eletrônico.

Os sites devem disponibilizar regras claras e características relevantes utilizadas para manter o usuário mais seguro. Como por exemplo, Políticas de Privacidade, capacidade de personalização, permissões para a manipulação dos dados coletados, consentimento prévio do

usuário, dentre outros itens que sejam capazes de analisar os sites através da inspeção durante a navegação.

Pensando nisso, definiu-se um cenário de estudo para análise de sites referente às características disponibilizadas aos usuários, sendo esses os sites de comércio eletrônico brasileiro com o ramo de atividade voltado ao varejo.

O varejo é definido como sendo a venda em pequenas quantidades (FERREIRA, 1989), sendo o comércio onde as mercadorias são vendidas por unidades, exercido por revendedores (os varejistas) que adquirem os bens dos produtores ou dos atacadistas.

A seguir é apresentada a metodologia empregada para o estudo de avaliação dos sites, definindo-se para isso algumas etapas a serem seguidas, de forma a conduzir a realização de experimentos.

Objetivou-se conhecer informações precisas sobre os itens a serem analisados nos sites, de forma que fossem úteis na comprovação da eficácia da utilização das taxonomias de privacidade e personalização. Dessa forma, com a verificação da utilização desses itens pelos sites, tornou-se possível quantificar o nível de privacidade e personalização oferecido pelos sites.

Essas informações são adquiridas através do conhecimento científico, que se dá pela investigação planejada, de acordo com normas guiadas por uma metodologia bem definida. Dessa forma, busca-se compreender ou explicar a realidade pela prática de um conjunto de procedimentos técnicos e intelectuais denominados métodos científicos (GIL, 2002).

## A.2. Metodologia Utilizada

De acordo com Meira Jr., Murta e Resende (2000), comércio eletrônico pode ser definido como o conjunto de técnicas e tecnologias computacionais utilizadas para facilitar e executar transações comerciais de bens e serviços físicos ou virtuais.

A escolha pelos sites de comércio eletrônico para o estudo deu-se ao fato de sua característica inerente de coleta de dados, devido à necessidade em identificar o consumidor e prover personalização ao mesmo. Essa coleta pode em vezes infringir a privacidade do usuário, sendo esse cenário um dos mais preocupantes durante a navegação pela web.

Para tanto, utilizou-se uma metodologia composta por um conjunto de métodos e macro-atividades complementares, de modo a ampliar o entendimento dos resultados na formação do conhecimento.

Tais métodos podem ser caracterizados como: i) indutivo, onde o conhecimento é fundamentado na experiência e indução para outros casos semelhantes e ii) dialético, baseado na interpretação dinâmica e totalizante da realidade, onde os fatos não podem ser considerados fora de um contexto social, político e econômico (LAKATOS e MARCONI, 2001).

Sabido do extenso número de sites existentes e da incapacidade humana de analisá-los, viu-se necessário utilizar uma metodologia de amostragem para escolha de quais sites seriam analisados individualmente.

A Figura 37 ilustra, de forma sintética as etapas gerais desta pesquisa, sendo que os detalhes dessas etapas, constituídos do projeto e desenvolvimento do Estudo de Caso, são apresentados nas próximas seções.


**Figura 37. Ilustração das etapas seguidas para o levantamento de dados**

As etapas compreendem um levantamento de dados para fornecer auxílio na identificação de características relevantes a serem analisadas na avaliação dos sites, para a classificação do nível de privacidade e personalização apresentado pelos mesmos.

Tais características podem subsidiar na determinação de diretivas de privacidade e personalização, a partir da taxonomia já proposta, podendo essas serem tomadas como chaves para a análise em outros trabalhos.

**Etapa 1 - Determinação dos sites, dentro do cenário de e-commerce brasileiro, a serem analisados.**

Como já mencionado, de acordo com métodos indutivos, é inviável a análise de todo o conjunto do cenário de estudo apresentado. Para suprir essa dificuldade, fez-se uma amostragem selecionando parcelas menores desse conjunto que representem o todo.

Para a amostra de sites selecionados foram levados em consideração alguns sites do comércio eletrônico brasileiro, apresentados por duas pesquisas feitas pela e-bit[25] e Info Exame[26]. Cada elemento da amostra foi analisado com a utilização de recursos de visualização presentes nos navegadores web.

A tarefa de observação possibilitou a predição de cada amostra em constituir-se de um exemplo positivo ou negativo, em relação à presença de características que seriam relevantes à análise e de avaliar a qualidade de seus conteúdos na composição do modelo.

**Etapa 2 - Determinação das características a serem analisadas.**

Após a determinação da amostra, nesse passo foram realizadas as atividades relacionadas à criação de modelos para determinação de quais características serão analisadas em cada site, sendo essas apresentadas sob a forma de itens.

Para a determinação desses itens foram levadas em consideração as taxonomias de privacidade e personalização desenvolvidas, as quais subsidiaram o desenvolvimento de itens que apresentam as características que são pertinentes aos sites apresentarem.

Tais itens foram dispostos em uma tabela, chamada Tabela Ocorrência, de modo que os sites possam ser analisados através da avaliação por inspeção manual e então assinalado na tabela caso o site apresente tal característica, ou seja, apresente o item que caracteriza alguma das camadas das taxonomias.

**Etapa 3 - Avaliação por inspeção para verificação da presença das características relevantes.**

Esta etapa é caracterizada pela realização da inspeção nos sites e posterior marcação na tabela. Se o site apresenta o item em questão, esse é assinalado, sendo também disponibilizada uma consideração sobre a análise. Os sites podem também não apresentar tais itens ou até mesmo, os itens não serem aplicáveis, dependendo do contexto.

Ao final da análise dos sites é apresentada uma tabela, retratando a porcentagem de sites que contemplam as camadas da taxonomia de privacidade e personalização, de modo que ao final do trabalho pode-se ser feita a comparação com a análise automatizada, com a utilização da "PrivPerson".

[25] http://www.ebit.com.br/ebit/html/index.asp
[26] http://info.abril.com.br - agosto de 2006

## A.3. Amostragem (Etapa 1)

Esta etapa envolveu a busca e a coleta de uma amostra do conjunto constituído pelo domínio dos sites de comércio eletrônico brasileiro, de modo a fazer observações nesses sobre as características apresentadas para verificação da utilização das camadas das taxonomias de privacidade e personalização.

De acordo com Mattar (2001) a amostragem é definida como sendo o processo de colher amostras de uma população, sendo a amostra um subconjunto da população total, que inclui todos os objetos dos quais ou sobre os quais pode-se coletar informações para atender os objetivos da pesquisa.

A amostragem probabilística, ou aleatória, é caracterizada pelo conhecimento da probabilidade de que cada elemento da população possa ser selecionado para compor a amostra. Essa probabilidade pode ou não ser igual para todos os elementos da população, entretanto precisa ser diferente de zero (MATTAR, 2001). Esse tipo de amostragem é subdividido em duas classificações menores, as quais são chamadas de amostragem probabilística simples e estratificada.

A amostragem probabilística simples é caracterizada pelo fato de que cada um dos elementos da população tem a probabilidade conhecida e idêntica à dos outros elementos, podendo qualquer um deles ser selecionado para fazer parte da amostra (MATTAR, 2001).

Se essa fosse utilizada, todos os sites teriam a mesma chance de serem selecionados para fazer parte da amostra. Como o número de sites é extenso, e alguns desses podem não ser considerados de conhecimento para uma parcela representativa de usuários que fazem compras pela Internet, poderia resultar em uma amostra não representativa.

Dessa forma, nesta pesquisa utilizou-se a amostragem probabilística estratificada e foi tomado como objeto da amostra sites de comércio eletrônico brasileiro, os quais representam de maneira adequada os demais sites. Assim, através da amostragem foi possível pesquisar apenas uma parte para inferir conhecimento sobre o todo, em vez de pesquisá-la completamente, o que nesse caso seria inviável devido ao grande numero de sites existentes.

Malhotra (2001) relata que o principal objetivo da amostragem probabilística estratificada é aumentar a precisão sem aumentar o custo. Por sua vez, Mattar (2001) diz que se eleva a precisão, a eficiência e a correção da amostra, de modo que essa seja uma amostragem significativa.

A seguir é descrito como a seleção dos sites foi feita. No entanto, informações referentes à identificação desses, como nomes, imagens e URLs não foram disponibilizadas,

devido ao fato de não ser o objetivo do estudo o enfoque em quais são os sites da amostra e sim, a relevância da análise e a contemplação das camadas das taxonomias por esses.

### A.3.1.Seleção de Sites para Análise

Por ser inviável a análise de todos os sites de comércio eletrônico, alguns desses foram selecionados para compor a amostra. Existem alguns sites de destaque no cenário de comércio eletrônico brasileiro, que certamente podem melhor representar os demais devido ao número de usuários que já os utilizam e às funcionalidades aplicadas aos mesmos, em decorrência do número de acessos e das necessidades impostas pelos usuários.

Em virtude disso, foi utilizada a amostragem probabilística estratificada, que tem como característica a divisão da população em estratos. Tais estratos podem ser definidos por: i) ranking de acesso e ii) tipo de venda, como varejo e atacado.

Para se determinar o tamanho da amostra, várias questões devem ser levadas em consideração, como questões estatísticas de representatividade e gerenciais.

Como apresentado por McDaniel e Gates (2003), a solução para a seleção não é o tamanho da amostra em relação ao tamanho da população e sim se a amostra é realmente capaz de representar a população.

Provas empíricas mostram que amostras pequenas, mas representativas, podem refletir com bastante precisão as características da população (MCDANIEL e GATES, 2003).

Como a amostragem escolhida é a probabilística estratificada, o primeiro passo após a determinação do tamanho da amostra é a divisão em estratos que representam a população. Dessa forma, foi delimitado um estrato dos sites de comércio eletrônico mais utilizados pelos usuários, sendo esses referentes ao varejo.

Para se ter conhecimento de quais seriam os elementos a constituir os estratos, utilizou-se pesquisas apresentadas pelos grupos: i) e-bit, onde são apresentados os 15 sites mais avaliados pelos usuários, estando esses intitulados como Diamante, Ouro e Prata (E-BIT, 2006) e ii) Info Exame, onde são retratadas as empresas que dominaram os negócios no país em 2005, sendo essas as maiores empresas de *e-commerce* no Brasil (REGGIANI, Ano 21).

Primeiramente é feita uma avaliação nos sites apresentados na pesquisa da e-bit e posteriormente nos sites apresentados pela pesquisa da Info Exame. Vale ressaltar que 4 dos sites analisados foram referenciados nas duas pesquisas, no entanto, sendo esses apresentados uma única vez nos resultados finais.

Para a preservação do anonimato das empresas foi considerado que o objetivo da análise é baseado na investigação de um cenário de comércio eletrônico e não em sites

individuais, tendo como resultado uma porcentagem de quantos sites apresentam as camadas das taxonomias.

Outra questão considerada no domínio da pesquisa é o descarte amostral inicial, sendo esse representado por sites que não estão ligados ao ramo de atividade varejo, mesmo que esses sejam referenciados na lista dos mais acessados. Tal descarte foi feito de modo a manter a proporcionalidade e a representatividade da amostra, sendo a amostragem constituída de 33 sites de comércio eletrônico brasileiro.

Os procedimentos adotados para os experimentos realizados, em parcelas menores de sites, são relativos ao planejamento e à execução do Estudo de Caso. Esses são descritos através de uma análise sobre a comprovação de quais questões são importantes observar para avaliação dos sites, sendo essas questões referentes as taxonomias.

## A.4. Tabela Ocorrência (Etapa 2)

Nesta seção é apresentada a marcação na tabela dos itens encontrados nos sites. Para a determinação de quais seriam os itens a serem considerados durante a inspeção nos sites, procurou-se utilizar descrições que representassem as camadas das taxonomias, enfatizando as características que são importantes serem apresentadas pelos sites.

Essas características são relevantes para facilitar a interação dos usuários com os sites, garantir a segurança oferecida e disponibilizar serviços de personalização com qualidade. Ainda são relevantes pois através delas tornou-se possível identificar as camadas das taxonomias nos sites analisados através da avaliação manual.

De acordo com a metodologia proposta, com as taxonomias de privacidade e personalização já apresentadas, foram identificados e enumerados 15 itens, os quais ocorrem ou deveriam ocorrer com freqüência nos sites.

Nesta análise não foi identificado como item a Camada 1 - Leis de Proteção de Privacidade da taxonomia de privacidade. Isso se deu devido à análise ser feita em sites de comércio eletrônico brasileiro e, como já apresentado na Seção 2.4, no Brasil não existe uma lei que defenda a privacidade do usuário durante sua interação com a web.

As descrições dos itens, bem como, a informação de quais camadas esses representam e uma consideração descrevendo-os pode ser vista na

Tabela 12. Cada item pode representar uma ou mais camadas, de acordo com a função ao qual foi atribuído.

**Tabela 12. Tabela Ocorrência dos itens identificados**

| Número Do Item | Camadas Contempladas | Item a ser analisado | Descrições sobre O item |
|---|---|---|---|
| 1 | Camada de Privacidade: 2 | Está definida uma Política de Privacidade. | Desenvolver e disponibilizar diretivas com a Política de Privacidade utilizada pela empresa. |
| 2 | Camada de Privacidade: 3 | Utiliza Certificação de Privacidade para garantia de segurança dos dados coletados e da política de privacidade definida. | Passar por certificação para garantir a segurança na utilização do site. |
| 3 | Camada de Privacidade: 4 | Possui mecanismo de Notificação. | Oferece ao usuário acesso a diversas informações sobre os perigos providos da navegação pela Internet. |
| 4 | Camada de Personalização: 2 | Utiliza Formulários para coleta de dados. | O site disponibiliza formulários que devem ser preenchidos pelo usuário durante sua navegação, para fins de personalização. |
| 5 | Camada de Personalização: 3 | Armazena informações na máquina do usuário (*cookies*). | Armazena informações na máquina do usuário, como *cookies*, a fim de prover personalização. |
| 6 | Camada de Privacidade: 4 | Informa o usuário que está armazenando informações em sua máquina (*cookies*). | Informa ao usuário que está sendo armazenado *cookies* em sua máquina, de forma que o mesmo possa se sentir seguro. |
| 7 | Camada de Privacidade: 6 | Possui ambiente seguro HTTPS. | Utiliza um ambiente seguro para transações, fazendo a encriptação de dados, autenticação de servidor, integridade de mensagem e autenticação de cliente. |
| 8 | Camada de Personalização: 6 | Oferece personalização ao usuário. | Permite que o usuário seja diferenciado através de técnicas de personalização. |
| 9 | Camada de Personalização: 1 | Não envia códigos maliciosos aos usuários, seja no momento de acesso ao site ou por e-mail. | O site não utiliza processos maliciosos para a coleta de informações dos usuários, para ter permissão sobre informações. |
| 10 | Camada de Personalização: 1 | Não envia e-mails sem autorização, com propagandas e outros conteúdos. | Não envia e-mail de propagandas e promoções do site, sem que o usuário autorize. |
| 11 | Camada de Privacidade: 5 | Possibilita que o usuário remova seu e-mail das listas de propagandas. | O site deve prover estruturas que permitam ao usuário selecionar o aceite ou não de seu e-mail nas listas de propagandas. |
| 12 | Camada de Privacidade: 4<br>Camada de Personalização: 5 | Informa quais dados estão sendo gravados. | Informa ao usuário quais de seus dados estão sendo gravados pelo site. |
| 13 | Camada de Privacidade: 5 | Permite que o usuário possa excluir seus dados pessoais. | O site possibilita que o usuário exclua suas informações caso julgue necessário. |
| 14 | Camada de | Permite que o usuário altere | Deve ser possível que o usuário tenha |

| | | | |
|---|---|---|---|
| | Privacidade: 5 | seus dados pessoais. | acesso a seus dados pessoais e possa alterá-los. |
| 15 | Camada de Privacidade: 6<br>Camada de Personalização: 4 | Não há cruzamento de banco de dados com repasse ou consulta de informações a terceiros. | O site não deve repassar os dados coletados na navegação dos usuários a terceiros, sem que o mesmo seja informado. |

Para a inspeção dos sites de comércio eletrônico brasileiro, foi tomado como estudo a Tabela Ocorrência representando os itens apresentados para cada site analisados. No entanto, os detalhes da execução desse estudo não foram mostrados, sendo apenas, ao final do capítulo, apresentada uma tabela contendo a porcentagem de sites que apresentam as camadas, de acordo com a marcação nessa tabela.

Os itens apresentados são algumas das características que devem ser observadas pelos projetistas de sites de comércio eletrônico, de modo a aproximá-los às características relevantes a satisfação do usuário.

A seguir, apresenta-se o modelo de marcação na Tabela Ocorrência quando o site está sendo avaliado, sendo a representação da ocorrência ou não do item assinalada da seguinte forma:

i) com ✓ se o item for apresentado pelo site;

ii) se estiver assinalado com ⊗ o site não apresentar tal item, e,

iii) com –, significa que o item não se aplica ao site. Como por exemplo, se não houver coleta de dados, o item 14 não será aplicado, pois trata de permitir que o usuário altere seus dados e, dessa forma, não se aplica, pois se não há coleta não há dados a serem alterados.

Para cada site e para cada item avaliado, uma consideração é apresentada, informando o porquê da marcação, como pode ser visualizado na Tabela 13.

**Tabela 13. Tabela de marcação dos itens analisados para cada site**

| Número | Item a ser analisado | Encontrado | Considerações |
|---|---|---|---|
| **1** | Descrição do Item 1 | ✓ | O site apresenta o Item mencionado... |
| **2** | Descrição do Item 2 | ⊗ | Não apresenta pois... |
| **...** | ... | ... | ... |
| **15** | Descrição do Item 15 | – | Não se aplica pois ... |

Como exemplo, pode-se observar que o item 1 encontra-se presente no site avaliado, ao contrário do item 2, ausente nesse caso, e por fim, sobre o item 15, vê-se que esse não se aplica à avaliação do site em questão.

Os detalhes da avaliação nos sites foram omitidos, devido à análise, exemplificação e explicação dos 33 sites serem extensas, sendo apenas mostrados, a seguir, de forma resumida, os resultados finais para a avaliação dos sites da pesquisa da e-bit e da Info Exame.

## A.5. Avaliação por Inspeção dos Sites (Etapa 3)

A inspeção dos sites, segue um caminho que inicia na verificação da sua estrutura como um todo, antes de se analisar as partes de interesse separadamente, aplicando assim o teste funcional como indicado em Pressman (2001) e Beizer (1990).

O teste funcional trata o software como uma caixa cujo conteúdo é desconhecido e só é possível visualizar o lado externo, podendo ser aplicados nos sites onde requisições são feitas pelo navegador (lado externo) e processadas por um servidor onde o usuário não tem acesso ao funcionamento interno. O funcionamento interno das estruturas de dados que compõem os sites, foram tratados como do tipo caixa preta, para cada necessidade de interação verificou-se apenas a resposta do site.

Esse método tem como vantagem o custo reduzido de aplicação já que foram aplicados apenas em alguns sites, que são considerados poucos perto da população existente, mas nem por isso deixam de ter um valor significante.

Deve-se mencionar que quanto à verificação de Política de Privacidade, considera-se como tal o texto que informa sobre as regras de privacidade dos dados dos usuários, não sendo necessário que esse esteja denominado como “Política de Privacidade”, já que não é seguido um padrão.

A personalização foi considerada como algo que diferenciasse o usuário, seja com a disponibilização de seu nome, tela de boas vindas mesmo que num estado inicial da personalização, ou até mesmo, com a sugestão de produtos adequados ao perfil do usuário, sendo esse em um estado mais avançado de personalização (LOBATO e ZORZO, 2007a).

A partir da avaliação por inspeção pôde-se observar algumas vantagens e desvantagens/problemas apresentados pelos sites, sendo alguns desses merecedores de destaque. A seguir são apresentadas as análises dos resultados encontrados na inspeção dos 33 sites de comércio eletrônico, área varejo.

### A.5.1. Análise dos Resultados dos Sites Apresentados pela e-bit

É possível notar que mesmo que alguns itens não estejam presentes em todos os sites, na maioria dos casos os itens foram marcados como tendo ocorrido no site, ✓, pois uma parcela representativa dos sites avaliados apresenta esses itens.

Isso indica que os itens realmente são relevantes à análise de sites e que as camadas de privacidade e personalização, utilizadas para embasamento à definição desses itens, são características importantes que um site de comércio eletrônico deve apresentar.

A Tabela 14 mostra de forma resumida a identificação da ocorrência dos itens analisados nos 15 sites da amostra, apresentados pela pesquisa da e-bit, dando enfoque principal na porcentagem do número de sites que apresentam os itens.

Para um melhor entendimento é mostrado o número de sites que apresenta o item, que não fazem uso de suas características e o número de sites em que o item não se aplica. Na próxima coluna, é apresentada apenas a porcentagem de sites que têm o item como uma de suas características.

**Tabela 14. Tabela Resumo/Porcentagem de ocorrência dos sites da e-bit**

| Número do Item. | Item Analisado | Número de sites por Marcação ✓ | ✗ | – | Porcentagem de sites que apresentam o item |
|---|---|---|---|---|---|
| 1 | Está definida uma Política de Privacidade. | 13 | 2 | 0 | 86,6% |
| 2 | Utiliza Certificação de Privacidade para garantia de segurança dos dados coletados e da Política de Privacidade definida. | 14 | 1 | 0 | 93,3% |
| 3 | Possui mecanismo de Notificação. | 8 | 7 | 0 | 53,3% |
| 4 | Utiliza Formulários para coleta de dados. | 15 | 0 | 0 | 100% |
| 5 | Armazena informações na máquina do usuário (cookies). | 15 | 0 | 0 | 100% |
| 6 | Informa o usuário que está armazenando informações em sua máquina (cookies). | 8 | 7 | 0 | 53,3% |
| 7 | Possui ambiente seguro HTTPS. | 15 | 0 | 0 | 100% |
| 8 | Oferece personalização ao usuário. | 8 | 7 | 0 | 53,3% |
| 9 | Não envia códigos maliciosos aos usuários, seja no momento de acesso ao site ou por e-mail. | 3 | 0 | 12 | 20% |
| 10 | Não envia e-mails sem autorização, com propagandas e outros conteúdos. | 14 | 0 | 1 | 93,3% |
| 11 | Possibilita que o usuário remova seu e-mail das listas de propagandas. | 12 | 2 | 1 | 80% |
| 12 | Informa quais dados estão sendo gravados. | 7 | 8 | 0 | 46,6% |
| 13 | Permite que o usuário possa excluir seus dados pessoais. | 1 | 14 | 0 | 6,6% |
| 14 | Permite que o usuário altere seus dados pessoais. | 13 | 1 | 1 | 86,6% |
| 15 | Não há cruzamento de banco de dados com repasse ou consulta de informações a terceiros. | 11 | 0 | 4 | 73,3% |

De acordo com a Tabela 14 observa-se, no item 1, que 86,6% dos sites definem uma Política de Privacidade, informando sobre as medidas de segurança que são oferecidas e sobre as regras impostas pelos sites.

Essas abordam as obrigações e deveres dos sites e dos usuários, como por exemplo o usuário se comprometer a disponibilizar informações verdadeiras e o site mantê-las seguras, de modo que não sejam passadas a terceiros. Apenas 2 dos sites avaliados não definem uma Política de Privacidade.

Para comprovar o cumprimento das diretivas declaradas na Política de Privacidade e para garantir a seriedade da empresa, utilizam-se certificados de segurança, presentes em 93,3% dos sites avaliados e tendo a *VeriSign* como a entidade certificadora mais utilizada.

Algumas empresas também possuem o selo Internet Segura, o qual foi criado a partir da união das maiores empresas da Internet. Esse selo tem por objetivo mostrar ao usuário que a empresa que o possui está cumprindo com as obrigações impostas em sua política de privacidade.

A Figura 38 mostra uma certificação digital com status expirado (*Expired*), o que pode trazer aborrecimento ao usuário e até engano, pois no site é mencionado que *"a segurança do XXX é certificado pela Verisign, no XXX suas compras com cartão de crédito e débito em conta corrente estão sempre seguras"*.


**Figura 38. Assinatura de segurança apresentada pelo site está com status expirado**

Mesmo com a validade do selo expirada, a empresa continua com o certificado na página inicial do site, o que traz ao usuário uma falsa sensação de segurança, já que nem todos têm a curiosidade de verificar a validade do mesmo, podendo sua privacidade estar sendo colocado em risco. Também foi observado erro ao se clicar no certificado, o que pode tornar para o usuário uma navegação frustrante e cheia de inseguranças.

No item 3 pode-se observar que nem todas as empresas estão preocupadas com a tranqüilidade do usuário, com sua segurança durante a navegação e com o esclarecimento de dúvidas dos mesmos. Apenas 53,3% da amostra de sites analisados contam com mecanismo de notificação, apresentado pela taxonomia de privacidade.

Tais mecanismos de notificação podem ser expostos na forma de janelas de aviso ou texto, disponíveis no site ou especificados em sua Política de Privacidade. Um exemplo de notificação observada é a informação ao usuário sobre a utilização de e-mails falsos, com propagandas enganosas, os quais utilizam o nome das empresas.

Sobre a medição da utilização dos artefatos de comércio eletrônico, pode-se observar que todos os sites utilizam a coleta das informações de forma explícita, através do uso de formulários, de modo a facilitar a navegação do usuário e prover aos mesmos serviços de personalização, como pode ser visualizado no item número 4.

Observa-se no item 5 que 100% dos sites armazenam informações na máquina dos usuários, como *cookies*, por medidas de personalização. No entanto, apenas 53,3% dos sites avaliados informam aos usuários sobre isso, sendo dessa forma armazenada informações sem o consentimento do mesmo, como pode ser visto no item 6.

Os *cookies* contêm uma identificação que permite ao site de comércio eletrônico criar um histórico das atividades de seus usuários, podendo os sites implementar os chamados cartões de fidelidade.

Por mais que os *cookies* não sejam mecanismos para invasão de privacidade, deve-se informar nas Políticas de Privacidade o porquê da utilização desses, quais os benefícios para a personalização do usuário, se com seu uso a segurança é prejudicada e como excluí-los, caso o usuário considere necessário.

Todos os sites utilizam ambiente seguro HTTPS para transações, item 7, seja a partir da página inicial do site ou apenas na hora da efetivação da compra, fazendo a encriptação de dados, autenticação de servidor e cliente e garantia da integridade das informações. No entanto, foi observado que em alguns sites é emitida pelo navegador uma mensagem, sob a forma de uma janela de aviso, informando ao usuário que existe um problema no certificado de segurança do site. Ao clicar em “Exibir certificado” foi verificado que o certificado foi emitido a um endereço diferente do endereço que o site está utilizando.

Também foram observados que, no acesso as páginas através do HTTPS, existem alguns elementos que não podem ser disponibilizados, mas continuam sendo referenciados. Para tanto, é exibida uma mensagem para o usuário, em forma de uma janela de alerta,

informando-o do que está acontecendo e deixando claro que existem alguns itens que não são seguros, mesmo que esses não causem dano algum.

O item 8 mostra que pouco mais da metade dos sites oferecem personalização aos usuários, estando entre os 53,3%. Essa pode ser uma personalização inicial, apenas mostrando o nome do usuário, ou mais avançada, com sugestões de produtos de acordo com o perfil do usuário, dando um tratamento diferenciado a cada um. Com isso, é proporcionada ao usuário uma menor busca de suas necessidades e uma maior satisfação, sendo esse um dos grandes pontos para aumento de vendas nos sites de *e-commerce*.

Por exemplo, foi observado que alguns sites não provêem nenhum tipo de personalização, disponibilizando apenas os produtos que estão na sacola de compras dos usuários. No entanto, o simples acesso à sacola de compras é considerado como requisito funcional e não um tipo de personalização.

Uma personalização bem diferenciada e já num estado mais elevado é a apresentada em apenas 1 site da amostra, onde os produtos são oferecidos ao usuário de acordo com seu perfil. Também são mostrados produtos já visualizados pelo usuário anteriormente e produtos comprados por outras pessoas que têm o perfil semelhante ao do usuário em questão.

O item 9 mostra que apenas 20% dos sites avaliados se preocupam em deixar claro que não enviam códigos maliciosos aos seus usuários, seja no momento do acesso ao site ou através do envio de e-mail, e que os usuários devem sempre estar atentos para o recebimento de e-mail maliciosos que usam o nome da empresa para chamar atenção do usuário.

Para evitar que o usuário seja enganado, alguns dos sites especificam quais são os endereços de e-mails que realmente são utilizados pelas empresas, para que os usuários possam excluir qualquer outro e-mail que pareça suspeito.

Para 15 dos sites avaliados, apenas 3 dizem não enviar códigos maliciosos aos usuários. Em 12 foi definido que esta opção não é aplicável, pois esses não mencionam sobre nenhum tipo de código malicioso ao qual o usuário deva estar atento e, durante o tempo da análise, não foi possível observar nenhum acontecimento que venha a comprometê-los.

No item 10, foi verificado que 93,3% dos sites não enviam e-mails de propagandas, promoções e outros conteúdos aos usuários sem sua prévia autorização. Para 1 site não foi possível analisar, pois não existe um cadastro prévio mas apenas um cadastro para a efetivação de compra, sem *login* e senha, estando esse enquadrado como não aplicável.

Dentre os sites da amostra, 12 desses permitem que os usuários removam seus e-mails das listas de propagandas caso não queiram mais recebê-los, tendo atingido os 80%. Dois dos

sites não permitem e em 1 é considerado não aplicável, pois não há um cadastro de recebimento de promoções e anúncios, como pode ser visualizado no item 11.

Para remover os e-mails dessas listas, alguns dos sites disponibilizam um endereço de e-mail para que o usuário possa enviar uma mensagem solicitando o cancelamento, outros permitem que o usuário edite seu cadastro e faça isto.

No item número 12, observou-se que apenas 7 dos sites avaliados, 46,6%, informam aos usuários quais são os dados coletados durante sua navegação. Esse item é de grande importância pois é relevante ao usuário saber quais de suas informações estão sendo coletadas, por motivos de garantia de privacidade, segurança e conseqüentemente, maior confiança no site.

Um item pouco utilizado pelos sites é o item 13, com 6,6% de ocorrência, o qual permite ao usuário excluir seus dados pessoais coletados. Muitos dos sites avaliados não têm interesse em disponibilizar tal opção, e nem esclarecem dúvidas dos usuários de como proceder.

É relevante às empresas manterem os dados de seus usuários, podendo utilizar essas informações apenas por interesses próprios, como na geração de estatísticas sobre os itens mais procurados e perfis de usuários.

Apesar de não haver a preocupação quanto à exclusão dos dados, o item 14, o qual trata da possibilidade do usuário alterar seus dados, está presente na maioria dos sites, 86,6%. Em apenas em 1 site essa opção não é possível e em outro não é aplicável, pois não existe um cadastro prévio e, conseqüentemente, não existe gerenciamento de conta.

Por fim, foi observado que 11 dos 15 sites avaliados disponibilizam informações, garantindo que não há cruzamento de banco de dados com repasse ou consulta de informações a terceiros. Dessa forma, os sites buscam oferecer maior segurança, pois as informações gravadas são confidenciais e de uso apenas da própria empresa para medidas de estatísticas e de personalização, estando esse item presente com uma porcentagem de 73,3% nos sites avaliados.

Após a observação do número de sites que apresentaram o uso dos itens, para alguns itens a análise foi feita de forma diferente com o objetivo de identificar se os sites que contemplam ou não os itens são sempre os mesmos.

Quanto ao item 1, pode-se verificar que apenas 2 dos sites não o apresentam e, no entanto, esses mesmos sites apresentam o item 2, o qual seria interessante ser utilizado em conjunto com o item 1.

Pode-se observar que apenas 1 dos sites que declaram uma Política de Privacidade não utiliza uma entidade certificando que tal política está sendo seguida.

Em relação ao item 3, foi constatado que 7 sites não possuem mecanismo de notificação, sendo que 6 desses não apresentam o item 12, informando ao usuário quais dados estão sendo gravados.

Pode-se observar que todos os sites utilizam formulários para coleta de informações dos usuários, item 4 e, no entanto, 7 sites não oferecem personalização ao usuário, item 8, não sendo dessa forma justificada a coleta.

Outro item também apresentado em unanimidade é o item 5, sendo feito por todos os sites a prática de armazenamento de *cookies* na máquina do usuário. No entanto, apenas 8 desses sites informam aos usuários sobre o armazenamento, item 6.

Dos sites analisados, 14 desses dizem não enviar e-mails sem autorização aos usuários, item 10. No entanto, 2 desses não possibilitam ao usuário remover o cadastro de seu e-mail das listas, item 11. Isto implica dizer que dois dos sites que dizem não enviar e-mails sem autorização, podem estar enviando, pois depois que o usuário aceitou recebê-los pela primeira vez não é possível voltarem atrás em suas escolhas.

Quanto ao item 13 e item 14 pode-se observar que apenas 1 site permite aos usuários excluírem seus dados e apenas um outro não permite alterar. Os sites que não permitem excluir os dados também não permitem alterar.

Com essa análise pode-se observar que a apresentação ou não dos itens pelos sites ocorrem dentre todos os sites da amostra, não sendo, por exemplo, apenas alguns sites específicos que não apresentam os itens.

### A.5.2. Análise dos Resultados dos Sites Apresentados pela Info

A avaliação para a análise dos resultados dos 22 sites apresentados pela Info Exame também foi feita baseada nas taxonomias de privacidade e personalização proposta, assim como apresentado na análise dos sites da e-bit.

Na Tabela 15 é mostrado, de forma resumida, o número de itens apresentados ou não pelos sites, e o número de sites em que o item não se aplica. Também é apresentada a porcentagem de sites que apresentaram os itens propostos para essa avaliação, marcados com o símbolo ✓ na Tabela Ocorrência.

**Tabela 15. Tabela Resumo/Porcentagem de ocorrência dos sites da Info**

| Número do Item. | Item Analisado | Número de sites por Marcação | Porcentagem de sites que apresentam o |
|---|---|---|---|

| | | ✓ | ✗ | – | item |
|---|---|---|---|---|---|
| 1 | Está definida uma Política de Privacidade. | 15 | 6 | 1 | 68,2% |
| 2 | Utiliza Certificação de Privacidade para garantia de segurança dos dados coletados e da Política de Privacidade definida. | 7 | 14 | 1 | 31,8% |
| 3 | Possui mecanismo de Notificação. | 5 | 16 | 1 | 22,7% |
| 4 | Utiliza Formulários para coleta de dados. | 21 | 0 | 1 | 95,4% |
| 5 | Armazena informações na máquina do usuário (cookies). | 17 | 4 | 1 | 77,3% |
| 6 | Informa o usuário que está armazenando informações em sua máquina (cookies). | 5 | 13 | 4 | 22,7% |
| 7 | Possui ambiente seguro HTTPS. | 18 | 3 | 1 | 81,8% |
| 8 | Oferece personalização ao usuário. | 12 | 9 | 1 | 54,5% |
| 9 | Não envia códigos maliciosos aos usuários, seja no momento de acesso ao site ou por e-mail. | 2 | 0 | 20 | 9,1% |
| 10 | Não envia e-mails sem autorização, com propagandas e outros conteúdos. | 16 | 5 | 1 | 72,7% |
| 11 | Possibilita que o usuário remova seu e-mail das listas de propagandas. | 14 | 7 | 1 | 63,6% |
| 12 | Informa quais dados estão sendo gravados. | 11 | 10 | 1 | 50% |
| 13 | Permite que o usuário possa excluir seus dados pessoais. | 4 | 17 | 1 | 18,2% |
| 14 | Permite que o usuário altere seus dados pessoais. | 18 | 2 | 2 | 81,8% |
| 15 | Não há cruzamento de banco de dados com repasse ou consulta de informações a terceiros. | 14 | 0 | 8 | 63,6% |

Analisando a Tabela 15 vê-se que nem todos os sites estão preocupados com a declaração de uma Política de Privacidade, onde são definidos quais são as diretivas impostas, bem como, quais são as medidas adotadas pelos sites para fornecer segurança e confiança ao usuário. Este item foi apresentado por 68% dos sites.

Muitas vezes, as informações que os usuários podem encontrar nos sites determinam suas ações na navegação e um dos pontos onde os sites podem conquistar os usuários é no esclarecimento de dúvidas e fornecendo informações relevantes em suas políticas.

A Figura 39 mostra um exemplo de uma certificação dada ao site pela *VeriSign* estando essa com o status revogado, mesmo estando o período de certificado válido, entre 07 de novembro de 2005 a 07 de novembro de 2006, sendo a data final o período em que tal estudo foi realizado.


**Figura 39. Exemplo de certificação revogada para o site**

Tal anulação não foi mencionada no site e ainda assim o site continuou declarando em sua Política de Privacidade que a segurança oferecida é definida pela certificação da *VeriSign*:

*"Nós da YYY entendemos sua preocupação com os dados fornecidos através de compras pela Web e por este motivo temos em nosso site além da assinatura VeriSign (empresa que emitiu nosso certificado de criptografia), nossa própria política, para que você não precise se preocupar e possa realizar suas compras com tranqüilidade"*.

Foram também encontradas certificações com o status expirado, onde a certificação venceu em 26 de fevereiro de 2003 e até hoje nenhuma medida havia sido tomada, estando o símbolo da Entidade Certificadora presente na página inicial do site.

Apenas 5 dos sites, 22%, apresentam mecanismos de notificação ao usuário sobre os perigos decorrentes da utilização da internet, medidas a serem tomadas para garantir maior segurança, dentre outras informações.

Já no item 6, apenas 5 sites informam ao usuário que estão sendo armazenandas informações em sua máquina, como por exemplo *cookies*. No entanto, como visto no item 5, 17 sites armazenam *cookies* na máquina do usuário, sendo que 12 desses armazenam informações sem o consentimento do usuário.

Quanto à utilização de ambiente seguro para a efetuação de transações, observou-se que existem sites que não utilizam o ambiente HTTPS nem no momento de acesso na parte da compra, onde tal serviço deveria ser oferecido de modo a trazer mais segurança ao usuário e maior confidencialidade pela empresa.

Um dos itens mais críticos é o item 9 onde apenas 2 dos 22 sites da pesquisa abordam assuntos referentes a códigos maliciosos, dizendo que nenhum tipo de arquivo executável é enviado ao usuário pela empresa, seja no momento de acesso ao site ou por e-mail. Os demais

sites não apresentam informações aos usuários. No entanto nada que pudesse comprometê-los foi observado durante a fase de análise, sendo esse item considerado não aplicado a esses sites.

Quanto ao item 13, 4 sites permitem que o usuário possa excluir seus dados, seja na forma de informações de como o usuário deva proceder para efetuar tal ação, ou seja na disponibilização de e-mails, ao qual o usuário poderá enviar um pedido de exclusão de sua conta cadastrada no site.

A seguir apresenta-se a análise enfatizando se os sites que apresentam ou não determinados itens estão sendo os mesmos para todos os itens, como foi feito na análise dos sites apresentados pela pesquisa da e-bit.

Sobre o item 1 em relação ao item 2, pode-se observar que 6 sites não oferecem Política de Privacidade, sendo que 5 desses também não utilizam certificação. No entanto, ainda 9 outros sites, os quais apresentam uma política, também não contam com uma entidade certificadora que garanta o cumprimento das regras definidas em suas políticas.

Em relação ao item 3, 16 dos sites não apresentam mecanismos de notificação, informando aos usuários sobre os perigos decorrentes da navegação pela web, sendo que 9 desses também não informam aos usuários sobre quais de seus dados são coletados durante a interação com o site, item 12.

O item 4 é contemplado por 21 dos sites, os quais utilizam formulários para a coleta de dados, estando apenas 12 desses disponibilizando serviços personalizados aos usuários, sejam esses serviços avançados e bem direcionados ou não, item 8.

Sobre o item 5, 17 sites armazenam informações na máquina dos usuários, sejam essas para beneficio dos usuários ou não. No entanto, apenas 5 desses informam aos usuários sobre esse armazenamento, item 6.

Dentre os 22 sites, 5 enviam e-mails sem a permissão dos usuários, item 10, e esses mesmos sites ainda não permitem que os usuários removam seus e-mails das listas, item 11. Há ainda outros 2 sites que permitem alteração, não permitindo a exclusão.

Quanto ao item 14, 18 sites permitem aos usuários alterarem suas informações cadastradas. No entanto, apenas 4 desses permitem aos usuários apagarem suas informações, item 13, sendo que 14 desses não fornecem tal opção.

## A.6 Conclusões

Todos os itens analisados durante a avaliação foram baseados em características que poderiam de alguma forma aumentar a segurança e a facilidade do usuário no acesso a serviços, principalmente no acesso a sites. De maneira geral, a presença desses itens pode converter-se em ganho de um maior número de usuários, que passam de visitantes a clientes e, com isso, gerar maiores lucros para a empresa.

A partir deste estudo de avaliação por inspeção nos sites de comércio eletrônico, foi possível validar o quão importante foi a definição das taxonomias de privacidade e personalização compostas pelas camadas, já que essas contemplam características que são disponibilizadas pelos sites.

Observa-se que muitos dos sites do comércio eletrônico brasileiro cumprem com as diretivas e apresentam características de relevância à navegação segura do usuário. No entanto, é preciso haver uma maior conscientização por parte das empresas sobre alguns tópicos que devem ser levados em consideração, pois a falta de informação pode causar insegurança no usuário, acarretando na desistência da navegação.

É importante, por exemplo, que os usuários possam ter controle sobre as informações que foram coletadas, estando aptos a editá-las ou até apagá-las, caso julguem necessário.

No Capítulo 6 são apresentadas tabelas, dando ênfase ao número de sites que apresentam as camadas das taxonomias de privacidade e personalização, sendo mostrada a porcentagem de sites da amostra que utilizam cada camada.

Através dessa pesquisa, também é possível ressaltar que algumas das empresas abusam da confiança dos usuários depositada nas entidades certificadoras, pois disponibilizam nos sites a imagem dos selos de segurança sem que a certificação esteja correta. Ao verificar a integridade da certificação atribuída aos sites, verificou-se que em alguns a certificação está expirada, revogada ou a aprovação do selo é para um endereço diferente do que o site está utilizando.

Um ponto positivo que se pôde observar quanto aos sites, é que uma parcela considerável da população amostral não envia e-mails sem autorização aos usuários, com propagandas e outros conteúdos, de forma a não causar aborrecimento quanto ao recebimento de e-mails indesejáveis, preservando em parte a privacidade dos usuários.

A maioria dos sites avaliados, tanto na pesquisa apresentada pela e-bit quanto pela Info, optaram pela certificação da *VeriSign*, no entanto podem ser observados outros tipos de certificação, como a certificação da *Thawte*[27].

Apesar dos sites contemplarem as camadas da taxonomia de personalização, apenas 1 site apresentou uma personalização mais detalhada ao usuário, disponibilizando produtos de acordo com seu perfil. No entanto, foram apresentados por alguns sites outros tipos de personalização, estando essas em um estado mais simples.

Dentre vários ganhos apresentados com essa pesquisa, um dos mais importantes é que pode-se verificar que as camadas de privacidade e personalização são utilizadas pelos sites.

Dessa forma, comprova-se a importância da existência dessas camadas e valida-se suas arquiteturas sob a forma de taxonomias, de modo que sirvam como a base sobre a qual os mecanismos podem apoiar-se para a definição e a oferta de seus serviços aos usuários. Conseqüentemente, essas taxonomias mostram-se suficientemente relevantes para a avaliação dos mecanismos de privacidade e personalização disponíveis, principalmente em um cenário de utilização em sites de comércio eletrônico.

Com o desenvolvimento deste estudo pode-se ainda observar que os sites não seguem um padrão para a definição e disponibilização de suas Políticas de Privacidade, o que pode prejudicar o entendimento dos usuários sobre as regras seguidas pelos sites e em vezes, impossibilitar o encontro dessas pelos usuários e por ferramentas automatizadas.

Baseado neste estudo foi definida uma padronização para a Política de Privacidade apresentada pelos sites. Tal padronização contempla características relevantes que devem ser disponibilizadas nas políticas e as quais foram analisados neste estudo, sob a forma de itens que representam as taxonomias. Tal padronização é mostrada no APÊNDICE B.

[27] http://www.thawte.com

# APÊNDICE B - Padrões para Desenvolvimento de Políticas de Privacidade

A partir de conclusões tomadas após a realização do Estudo de Caso, foi possível observar que a maioria dos sites não define uma Política de Privacidade com regras e deveres claros aos usuários e quando as fazem, não seguem uma padronização.

Uma pesquisa publicada por Turow (2003) mostra alguns dados sobre a relação dos usuários com as Políticas de Privacidade, ressaltando a insatisfação, a falta de compreensão, e a necessidade de informação dos mesmos quanto às políticas disponibilizadas.

Já um trabalho realizado por uma equipe de pesquisadores da *North Carolina State University* identificou que dentre 40 Políticas de Privacidade examinadas, 12 requeriam para seu entendimento um nível de escolaridade superior e 7 requeriam o equivalente ao nível de pós-graduação (ANNIE *et al.*, 2004).

Com base nesses estudos, ressalta-se que tais políticas deveriam informar aos usuários sobre o que é feito para garantia da privacidade dos mesmos e quais técnicas são utilizadas para prover personalização. Também deveriam informar sobre a manipulação dos dados coletados, utilização de entidades certificadoras, armazenamento de informações na máquina do usuário, dentre outras questões que abordam a privacidade, segurança e personalização.

Preocupados com isso, é proposta uma padronização para as Políticas de Privacidade a serem disponibilizadas pelos sites, de modo a tentar aproximá-las ao entendimento do usuário e englobar todos os pontos interessantes a serem ressaltados em uma política escrita de maneira objetiva e clara. Esses pontos, ou Padrões foram criados de acordo com as características apresentadas pelas taxonomias, Capítulo 5, e às quais foram dispostas em itens para a análise dos sites, APÊNDICE A.

Para a padronização foram utilizados Padrões (*Patterns*), que, de acordo com Borches (2000), podem ser entendidos como uma forma de expressar conhecimento por meio de textos e esboços em um formato estruturado, cuja solução é de sucesso já que os mesmos podem ser utilizados e aplicados a outros problemas, os quais ocorrem frequentemente em um determinado contexto.

O interesse emergente em Padrões representa um esforço tanto para organizar e disseminar temas quanto para oferecer uma documentação de soluções comprovadas para problemas comuns.

Os Padrões de Política de Privacidade, desenvolvidos neste trabalho, abordam principalmente assuntos referentes à segurança, privacidade e coleta de dados dos usuários.

Na apresentação dos Padrões foram omitidos os detalhes de seu desenvolvimento e da explicação detalhada de cada um, seguidos dos exemplos apresentados. No entanto, todas essas informações podem ser verificadas através do estudo apresentado por Lobato e Zorzo (2007c).

## B.1. Padrões e suas Principais Características

Os Padrões são utilizados em várias abordagens e definidos por diferentes autores em suas respectivas áreas de atuação, no entanto todas as definições mostram um principal objetivo para os Padrões: o seu reuso (LOBATO e ZORZO, 2007b).

Para que determinadas regras venham a ser um Padrão, essas devem apresentar usos conhecidos que comprovem sua eficácia na resolução do problema (OLIVEIRA, BALBY e GIRARDI, 2004).

Dessa forma, o objetivo dos Padrões é apresentar um problema e propor a solução, o contexto no qual o problema ocorre, restrições que levam a diferentes considerações durante a elaboração da solução e as conseqüências de uso do Padrão, documentando a essência da solução, para que essa possa ser utilizada inúmeras vezes em situações similares.

Utilizando-se Padrões, neste caso, de Política de Privacidade, há um ganho no desenvolvimento das políticas dos sites, já que as soluções, definições e diretivas já encontram-se desenvolvidas e comprovadas suas relevâncias. Assim, esses poderão ser seguidos, proporcionando um ganho no tempo de busca e entendimento por parte dos usuários, já que esses saberão onde tais políticas encontram-se e os assuntos estarão dispostos de forma clara ao seu entendimento (LOBATO e ZORZO, 2007b).

Além disso, a comunicação entre desenvolvedores é facilitada, pois, falando-se em termos de Padrões pode-se elevar o grau de abstração durante a discussão da solução (SILVA *et al.*, 2005).

Com a reutilização dos Padrões, melhora-se a qualidade das políticas desenvolvidas pelos sites, já que as soluções foram elaboradas em vista a um problema recorrente, comprovadas e testadas, diminuindo os riscos de desenvolver Políticas de Privacidade que não sejam claras e satisfatórias às necessidades dos usuários.

## B.2. Metodologia para Formalização de Padrões

A partir dos passos seguidos no Estudo de Caso, foram observados que algumas camadas das taxonomias de privacidade e personalização não tem sido encontradas em todos os sites da amostra.

Baseado nessa observação, para a definição dos Padrões foram utilizados algumas dessas camadas de modo a definir uma Política de Privacidade que apresente regras claras e sensatas aos sites e atenda às necessidades dos usuários.

O formato e estilo de escrita dos Padrões foram baseados na "Linguagem de Padrões para escrita de Padrões" de Meszaros e Doble (1996) , onde é especificado que os Padrões são mais fáceis de compreender e aplicar quando alguns elementos estão presentes no formato utilizado, como:

- **Nome (numeração)**: permite uma referência rápida e comunica a idéia principal do Padrão. Pode-se utilizar uma numeração para facilitar a localização do mesmo;
- **Contexto**: descreve o problema encontrado para se ter a necessidade de padronização e a solução implantada;
- **Problema**: apresenta a problemática a qual o Padrão se aplica;
- **Forças**: informa os aspectos que influenciam a utilização do Padrão;
- **Solução**: apresenta a mensagem para a solução do Problema;
- **Conseqüências**: aborda os resultados decorrentes da aplicação da solução;
- **Usos Conhecidos**: mostra exemplos bem reconhecidos da aplicação prática do Padrão.

Neste trabalho, além dos elementos base para a definição dos Padrões, também utilizou-se o elemento Padrões Relacionados, o qual foi considerado importante para o entendimento dos Padrões formalizados. Esses devem ser nomes de outros Padrões que tenham alguma relação de contexto com os desenvolvidos.

Com a utilização desses Padrões a satisfação dos usuários, durante a navegação pela web tende a ser maior, já que terão informações bem definidas nas Políticas de Privacidade sobre os serviços oferecidos pelo site e sua segurança.

### B.2.1.Diretivas para Embasamento aos Padões

Para a definição dos Padrões, além da utilização das taxonomias de privacidade e personalização e das características observadas no Estudo de Caso, também foram considerados alguns princípios impostos por duas organizações.

Esses princípios foram considerados de forma a definir um escopo de uma solução de sucesso que deva ser seguido, facilitando o reuso para os demais projetistas, de maneira a propor e efetivar tais Padrões.

Como já mencionado na seção 2.4.2, existem organizações que destacam-se no cenário internacional, as quais tem por objetivo regularizar a proteção de privacidade dos usuários da web, sendo algumas delas: *Organization for Economic Co-operation and Development* (OECD) e da *Federal Trade Commission* (FTC).

Os princípios estabelecidos pela OECD, tratam da proteção de privacidade dos usuários, disponibilizando e retratando documentos específicos para a segurança, ainda especificam de que forma as informações pessoais dos usuários devem ser protegidas, sendo eles: Princípio do Limite de Coleta; Princípio da Qualidade dos Dados; Princípio da Especificação de Objetivo; Princípio da Limitação de Uso; Princípio da Segurança; Princípio da Transparência; Princípio da Participação Individual; Princípio da Responsabilidade.

Pela FTC são definidos alguns princípios de Práticas Justas de Privacidade, os quais são uma sintetização dos 8 princípios apresentados pela OECD. Esses princípios são baseados e desenvolvidos sob uma legislação para as práticas recomendadas de privacidade que protegem as informações pessoais de serem coletadas e mantidas pelo governo americano, sendo eles: Notificação; Escolha; Acesso; Segurança.

Dessa forma os princípios abordados por essas organizações junto às camadas das taxonomias de privacidade e personalização, foram utilizados como medidas para a definição dos Padrões de Políticas de Privacidade.

A seguir são apresentados os Padrões definidos, seguidos dos objetivos de seu desenvolvimento.

## B.3. Coleção de Padrões Definidos

Nessa seção são apresentados os Padrões definidos para Política de Privacidade, no entanto, os elementos que representam a descrição desses, incluindo os problemas observados, a motivação encontrada para sua definição e os exemplos, foram omitidos.

Os Padrões definidos são organizados hierarquicamente em uma coleção, formando a base para uma futura formalização em Linguagem de Padrões. São apresentados agrupados por níveis de abstração, do nível 1 ao 4, e retratados através de nós, no modo por largura na estrutura de árvore e analisados da esquerda para direita, seguindo a numeração atribuída.

Cada nó apresenta uma particularidade especial, sendo que apenas o conjunto desses dão sentido aos Padrões desenvolvidos. A partir da raiz, foi utilizada uma distância de um nó para a apresentação dos Padrões relacionados, como pode ser visualizado na Figura 40.

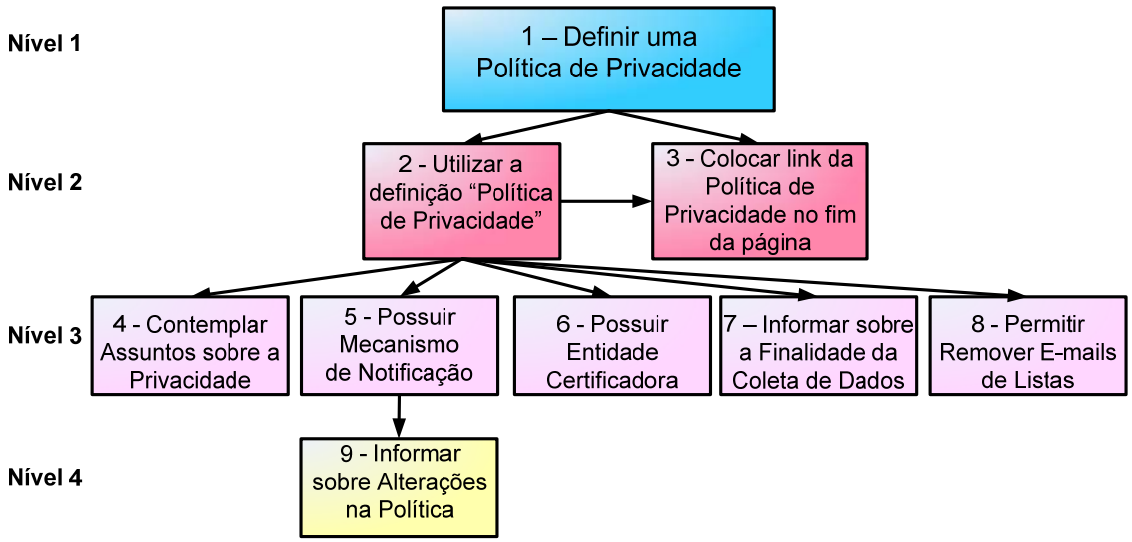


**Figura 40. Definição dos Padrões para Política de Privacidade**

É recomendado utilizar os Padrões em conjunto, observando também o uso com outros já desenvolvidos para a web, de forma a criar uma arquitetura de soluções eficientes e testadas, já que os Padrões representam a solução para um problema recorrente.

## B.4. Conclusões e Aplicação dos Padrões

Após apresentar os Padrões, os quais são a estrutura base para o desenvolvimento de Políticas de Privacidade, é possível aplicá-los observando os requisitos desejados para que tal Política seja disposta de forma clara, explicativa e com itens de interesse dos usuários.

A seguir é apresentada, como resultado final, uma aplicação prática de uma Política de Privacidade. Essa exemplificação foi desenvolvida de modo a ressaltar os requisitos desejados e analisando e as informações que se deve disponibilizar para desenvolver e manter a Política de Privacidade mais aproximada à uma política ideal.

**Tabela 16. Exemplo de Política de Privacidade**

**Política de Privacidade**

**Atualizada em 28/03/2007.**

**Sobre esta Política de Privacidade**
Esta Política de Privacidade foi estabelecida para o *"site Exemplo"* com o objetivo de assegurar a confiança e o sigilo das informações coletadas dos usuários.

Sabemos o quanto é importante para você conhecer e estar seguro sobre a utilização dos seus dados pessoais. Por isso, nos preocupamos em esclarecer e divulgar nossa política de utilização dessas informações. Assim, você poderá entender melhor quais informações obtemos e como as utilizamos.

**Dados Coletados**
Solicitamos informações quando você:

- Se cadastra no site (para agilização do processo de compra e para fins de estatísticas);
- Efetiva um pedido;
- Responde uma pesquisa on line;
- Participa de uma promoção;
- Cadastra-se em nosso boletim eletrônico ("*mail list*").

De forma automatizada, os seguintes dados também são coletados:

- Endereço IP;
- Data e horário do acesso;
- Tempo de leitura de cada página;
- Seqüência de páginas visitadas;

**Cadastro**
Não é necessário fornecer informações pessoais para navegar no site. Entretanto, para utilizar alguns dos serviços, será necessário identificar-se, fornecendo previamente alguns dados de caráter pessoal.

As informações serão armazenadas em um servidor seguro, e não são compartilhadas com terceiros.

**Finalidade da Coleta**
Inicialmente os dados coletados terão fins estatísticos. Para analisar, por exemplo, a quantidade de usuários que leram a Política de Privacidade, as diferenças entre as preferências de privacidade dos usuários e a freqüência de visita a cada página.

Os dados coletados serão também analisados para obter algumas informações sobre o perfil dos usuários que acessam o site, de modo a oferecer serviços personalizados.

Ainda utilizamos as informações coletadas por motivos de fins estatísticos, para efetivação da compra, andamento das operações e entrega de produtos.

**Exclusão das Informações**
O site possibilita que o usuário exclua e edite suas informações cadastradas no site, caso julgue necessário.

**Segurança**
Todos os dados coletados são armazenados em servidores internos e seguros, em um banco de dados reservado e com acesso restrito ao administrador deste site. Dessa forma, a manipulação dos dados se dá de maneira automatizada, não permitindo que pessoas não autorizadas tenham acesso aos mesmos.

**Certificação**
As práticas efetuadas pelo site seguem as diretivas definidas nessa política e são certificadas por

uma Entidade Certificadora, chamada *XXX*, a qual garante que a Política de Privacidade está sendo seguida.

Confira o certificado de segurança clicando aqui.

**Ambiente para Transações**
Utilizamos um ambiente seguro para transações, fazendo a encriptação de dados, autenticação de servidor, integridade de mensagem e autenticação de cliente.

**Tenha Cautela**
É possível que nossas páginas contenham *hiperlinks* que o levem a sites de terceiros. Recomendamos a leitura da Política de Privacidade desses sites, uma vez que não temos nenhuma responsabilidade sobre os mesmos.

Algumas pessoas utilizam do nome de empresas de responsabilidade para enviar e-mails aos usuários e também podem ser enviados juntos a esses e-mails códigos executáveis. No entanto, em hipótese alguma, os aceite, pois tais e-mails e executáveis tem o intuito de coletar suas informações pessoais.

Esteja atento a esses e-mails, prestando atenção no endereço do remetente, e, se possível entre em contato conosco avisando sobre o ocorrido.

**Envio de E-mails**
Este site não envia e-mail de propagandas e promoções do site sem a autorização do usuário.

O site provê estruturas que permitem ao usuário selecionar o aceite ou não de seu e-mail nas listas de propagandas. Assim, você poderá cancelar o envio de e-mails a qualquer momento.

**Cookies**
*Cookies* são pequenos arquivos de texto enviados ao seu computador e que são armazenados no mesmo. Estes arquivos servem para reconhecer, acompanhar e armazenar a navegação do usuário na Internet.

O uso de *cookies* possibilita ao site oferecer um serviço mais personalizado, de acordo com as características e interesses dos usuários, possibilitando, inclusive, a oferta de conteúdo e publicidade específicos para cada um.

**Alterações nesta Política**
Para assegurar regras claras e precisas, podemos eventualmente alterar essa política, e sendo assim, recomendamos sua leitura periodicamente.
Qualquer alteração na Política de Privacidade será transcrita na mesma.

No início da Política de Privacidade é indicada a data da última alteração, para facilitar ao usuário saber quando houve modificações.

**Considerações Finais**
Em caso de alguma divergência sobre nossa Política de Privacidade ou reclamações sobre os serviços prestados, sinta-se livre para entrar em contato conosco:

*Nome Fantasia da Empresa ou Site*
*Nome de registro no CNPJ da Empresa*
*Endereço físico*

Atendimento telefônico:
(0xxXX) XXXX.XXXX das XX:XXhs as XX:XXhs

Atendimento eletrônico:
http://www.empresa.com.br/antendimento
atendimento@empresa.com.br

Como já mencionado, esse modelo de Política de Privacidade foi desenvolvido baseado nos Padrões apresentados. Dessa forma o uso desses se torna viável durante a

elaboração e construção de Políticas de Privacidade que apresente características relevantes aos usuários e que atendem às verdadeiras exigências que uma política deve apresentar, sendo essa uma política de sucesso.

Tal modelo de política pode ser utilizado de modo a trazer facilidades aos sites na definição de suas Políticas de Privacidade e principalmente, trazendo benefícios aos usuários. Assim, essas serão definidas de maneira mais clara e objetiva, disponibilizada em uma linguagem que o usuário entenda, de modo a aumentar sua satisfação na interação com o site, já que o mesmo se sentirá mais seguro. Ainda é referenciada por um nome sugestivo, “Política de Privacidade” e de fácil localização.

# APÊNDICE C - Arquitetura da “PrivPerson”

A modelagem da ferramenta “PrivPerson” foi documentada na linguagem de modelagem *Unified Modeling Language* (UML).

A UML é uma linguagem gráfica para visualização, especificação, construção e documentação de artefatos de sistemas de softwares. É utilizada também para modelagem de negócio e outros sistemas que não são softwares, representando uma coleção das melhores práticas de engenharia na modelagem de sistemas (OMG, 2006).

Como mencionado por Booch, Rumbaugh e Jacobson (2005) a UML proporciona uma forma-padrão para a preparação de planos de arquitetura de projetos de sistemas. Essa inclui aspectos conceituais tais como, processos de negócios e funções do sistema, além de itens concretos como as classes escritas em linguagens de programação, esquemas de banco de dados e componentes de software reutilizáveis.

A seguir são apresentados os diagramas utilizados para a documentação da ferramenta desenvolvida, mostrando suas funcionalidades, atividades e detalhes de desenvolvimento.

## C.1. Diagramas de Representatividade

Um diagrama de atividade é essencialmente um fluxograma, sendo esse responsável por dar ênfase às atividades que ocorrem ao longo do tempo, constituindo basicamente de um nome e um conteúdo gráfico que são a projeção em um modelo.

Na Figura 41 pode ser visualizado o diagrama de atividades, mostrando o fluxo de controle de uma atividade para outra, quais foram os passos seguidos durante a tomada de decisões de qual caminho seguir para o desenvolvimento da “PrivPerson”.

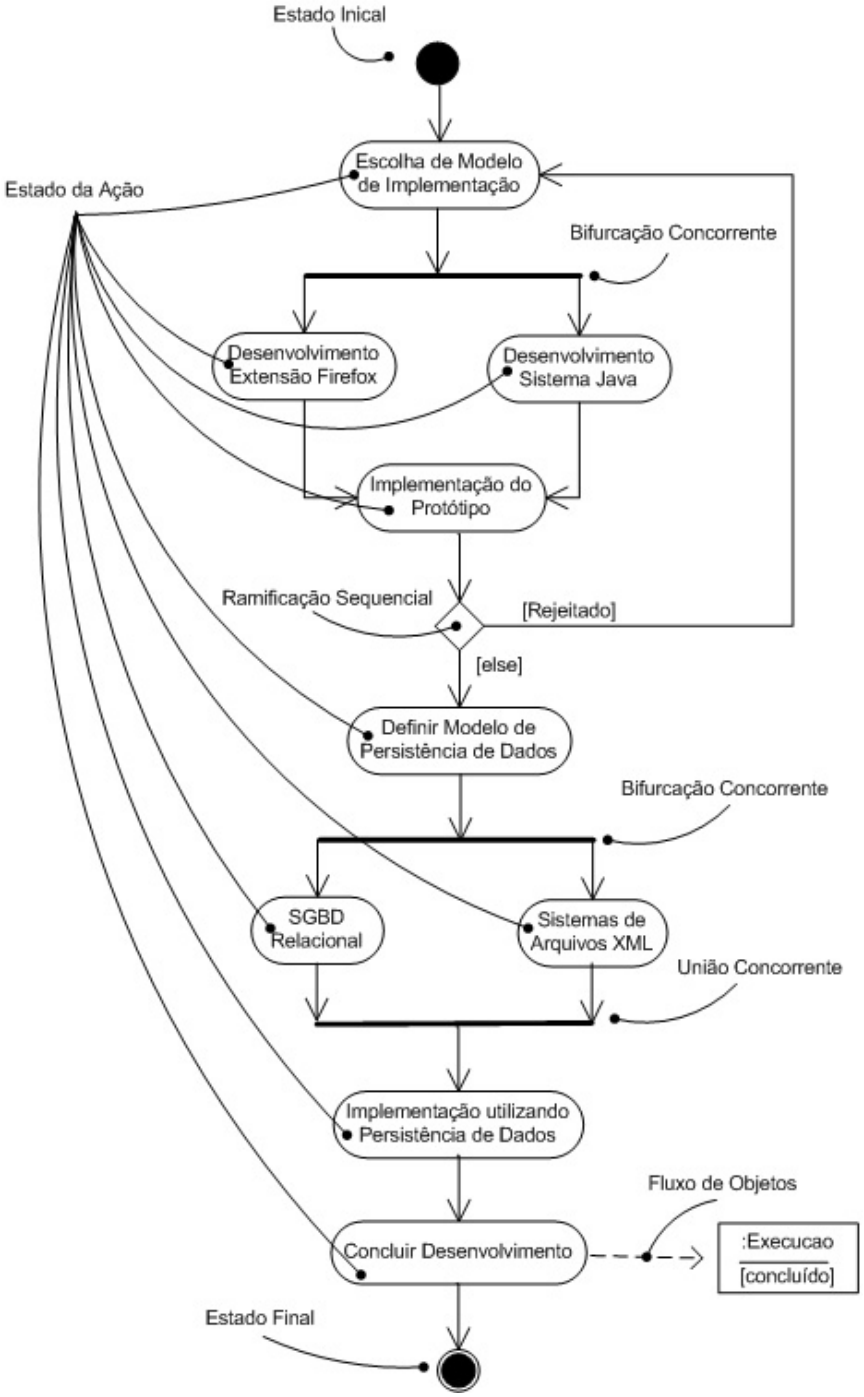


**Figura 41. Diagrama de Atividades**

Embora essa seja uma simplificação de todas as dúvidas encontradas e decisões tomadas quanto ao desenvolvimento da "PrivPerson", essa representação mostra de forma sintética passos importantes e capta o percurso crítico do fluxo de trabalho correspondente ao que foi desenvolvido.

A seguir, na Figura 46 apresenta-se o Diagrama de Caso de Uso. Esse ilustra uma visão global da "PrivPerson", apresentando um conjunto de opções que podem ser executadas, ou seja, os casos de uso, os atores e seus possíveis relacionamentos.

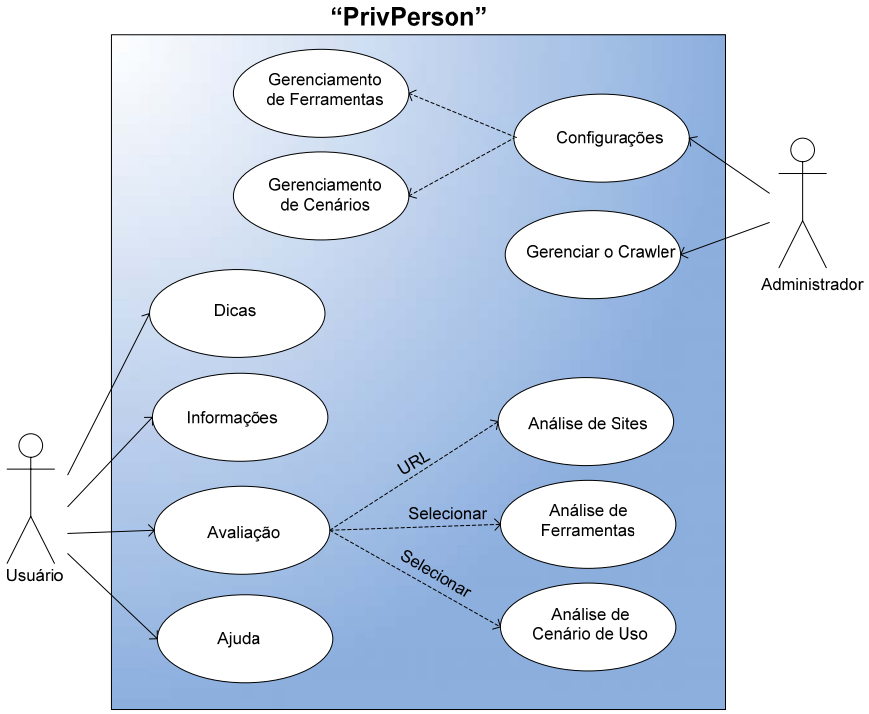


**Figura 42. Diagrama de Caso de Uso da "PrivPerson"**

De acordo com Booch, Rumbaugh e Jacobson (2005) esse tipo de modelagem é importante para a organização dos comportamentos envolvidos em situações.

Nesse diagrama pode-se verificar a modelagem da visão estática do caso de uso da "PrivPerson". Essa visão proporciona suporte principalmente para o comportamento de um sistema, sendo esses os serviços externamente visíveis que a "PrivPerson" fornece no contexto de seu ambiente.

Ao fazer a modelagem da visão estática de caso de uso de um sistema, pode-se aplicar o diagrama de caso de uso em uma de duas maneiras: i) fazer a modelagem do contexto do sistema, que foi a mostrada na Figura 42 e ii) fazer a modelagem dos requisitos de um sistema, que envolve a especificação do que o sistema deverá fazer independente de como deverá fazê-lo, sem mostrar ligação entre os autores e seus possíveis relacionamentos.

Uma próxima visualização que é considerada importante para o detalhamento das funcionalidades da "PrivPerson", é composta de um Diagrama de Classes contendo os pacotes principais de sua arquitetura, visualizado na Figura 37.

O Diagrama de Classes mostra um conjunto de classes, interfaces e colaborações com seus relacionamentos. São importantes não só para a visualização, a especificação e a documentação de modelos estruturais, mas também para a construção de sistemas executáveis (BOOCH, RUMBAUGH e JACOBSON, 2005).

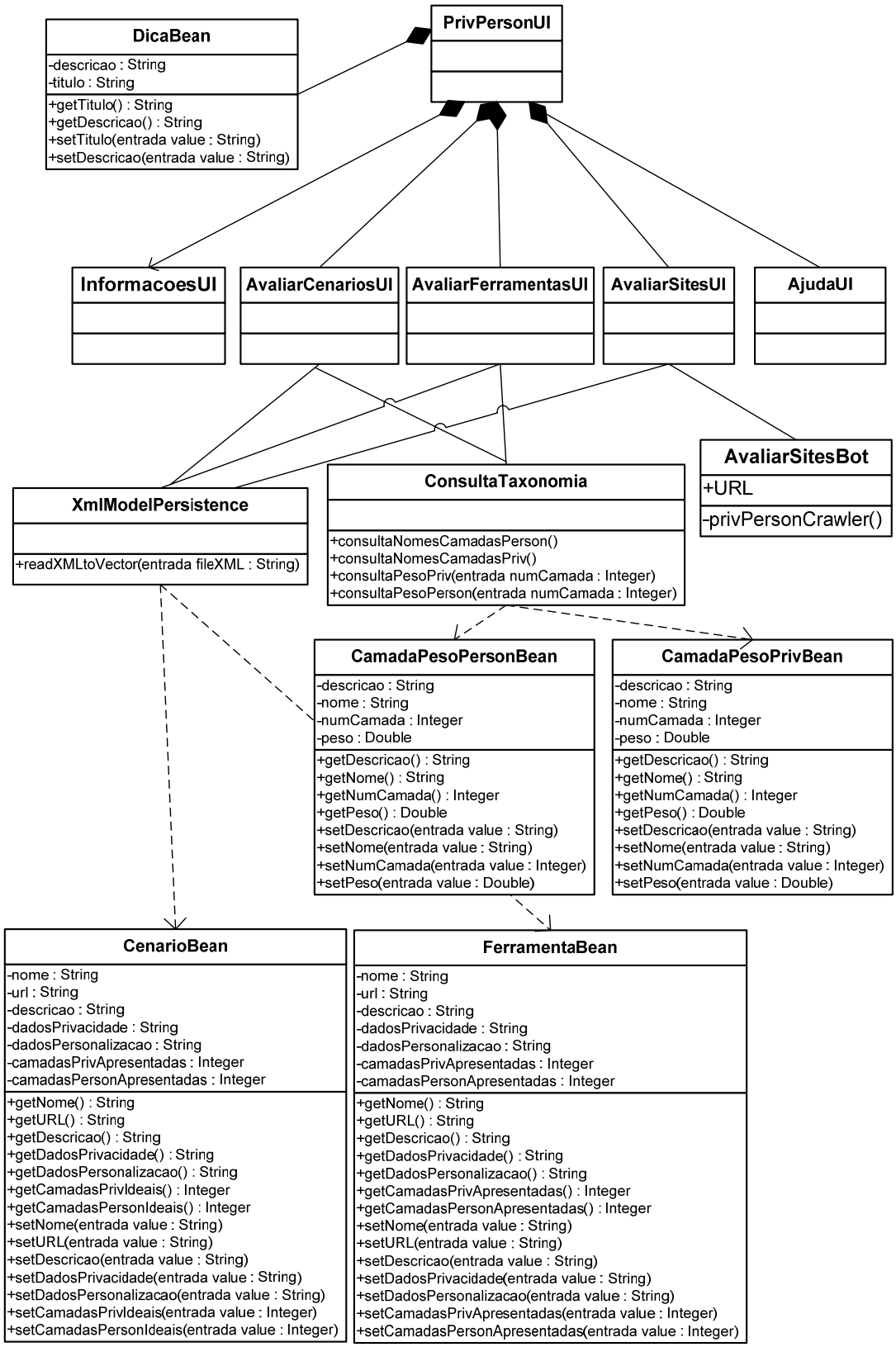


**Figura 43. Diagrama de Classes da modelagem da "PrivPerson", parte dos usuários**

Nessa figura são apresentadas as principais classes criadas para o funcionamento da "PrivPerson", as quais abordam as funcionalidades relevantes e ações desenvolvidas e disponibilizadas para a utilização dos usuários.

No entanto, as ações disponibilizadas pelos usuários também podem ser efetuadas pelos administradores, sendo dessa forma a Figura 43 constituinte de funcionalidades dos administradores também.

A seguir, na Figura 44 é mostrado outro Diagrama de Classes, contendo as classes que foram desenvolvidas, especificamente para compor as ações que os administradores podem efetuar na "PrivPerson".

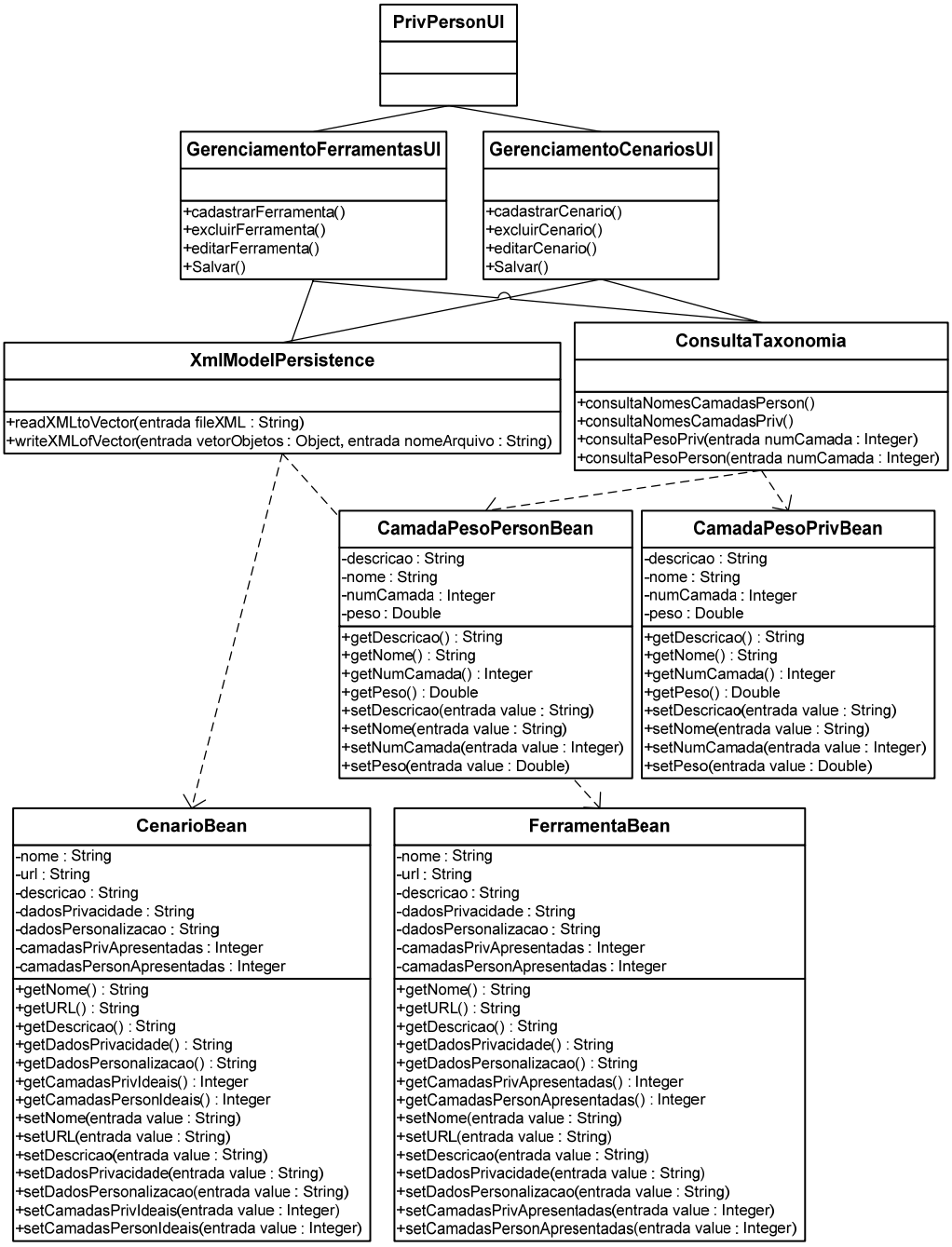


**Figura 44. Diagrama de Classes da modelagem da "PrivPerson", parte do admin**

Tais ações são de uso exclusivo dos administradores por representar o gerenciamento das ferramentas e cenários de utilização apresentados pela “PrivPerson”, podendo cadastrá-los, editá-los e excluí-los.

Os aspectos físicos da “PrivPerson” são apresentados de forma sintética na Figura 45, compondo do Diagrama de *Deployment* ou de Implantação, o qual demonstra a topologia da “PrivPerson”, possibilitando a visualização da distribuição física de seus componentes.

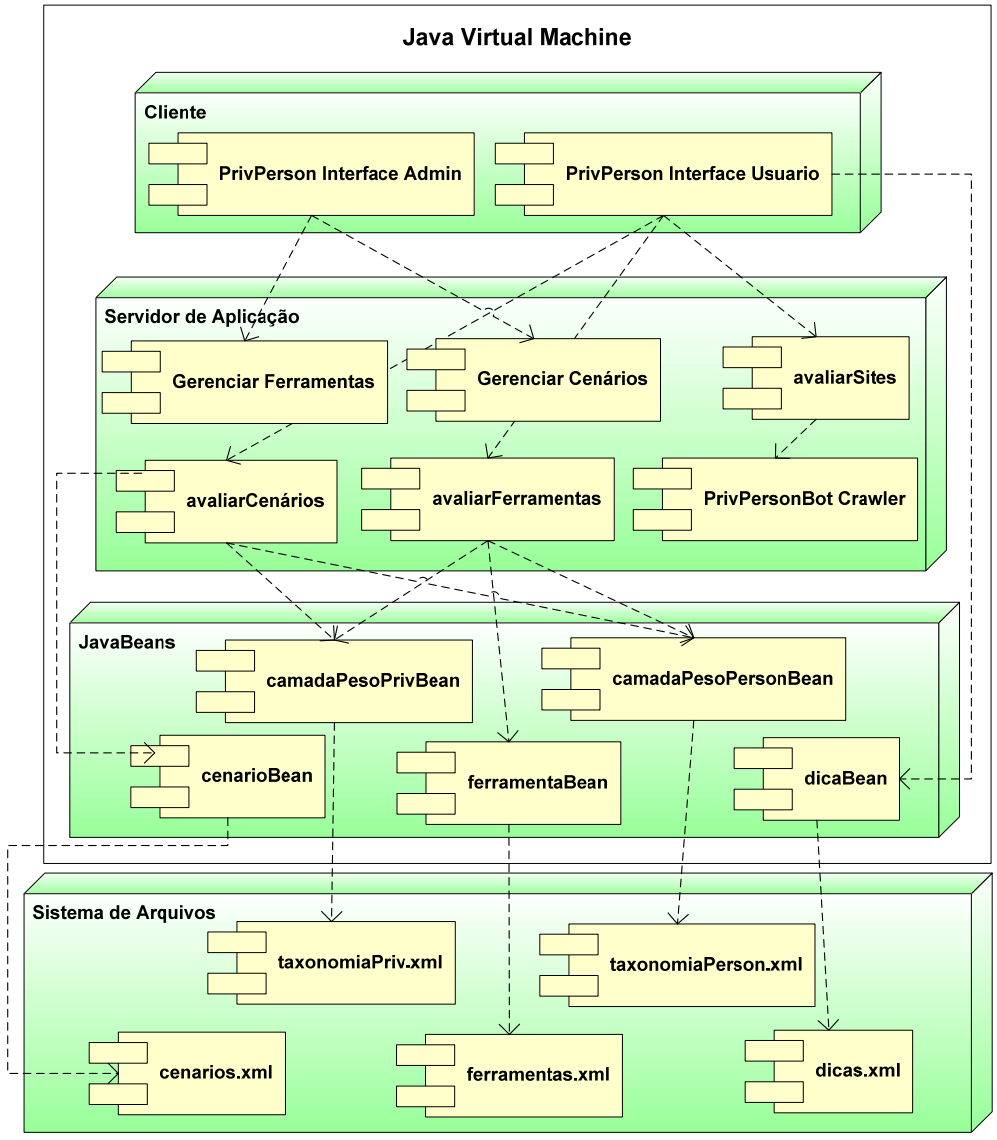


**Figura 45. Diagrama de Deployment ou de Implantação**

O Diagrama de *Deployment* é constituído de componentes de software, nós, conexões entre nós, objetos físicos e componentes que, juntos, compõem toda a arquitetura física da “PrivPerson”.

Nesse diagrama podem ser vistos alguns detalhes de implementação, assim como os relacionamentos das principais classes da “PrivPerson”, apesar de serem mostrados de forma simplificada.

## C.2. Implementação

Para a codificação das classes que constituem a solução descrita, ou seja, a "PrivPerson", utilizou-se a linguagem Java, devido a diversos motivos, tais como:

- Compatibilidade da linguagem com o paradigma de programação orientada a objetos;
- Disponibilidade de ferramentas de edição, compilação e execuções gratuítas;
- Existência de padrões adequados para documentação do código (JavaDoc);
- Portabilidade e reutilização do código;
- Disponibilidade de recursos para otimização de operações e para chamadas de programas executáveis externos, fornecidos com a API NetBeans.

Além de atributos, métodos e construtores pertinentes de todas as classes Java, utilizou-se a abordagem de componentes *beans* para facilitar a implementação.

Os *beans* podem ser definidos como sendo componentes de software que permitem a geração de partes reutilizáveis, ou seja, uma vez criados, podem ser reusados sem modificação de código pelas classes do pacote ao qual pertence. Os *beans* ainda apresentam as seguintes características:

- Propriedades, que são atributos que podem ser configurados para mudar a aparência ou o comportamento de um *bean*, como exemplo, botão habilitado ou não;
- Introspecção, capacidade de oferecer informações sobre si mesmo. O *bean* possui estratégias para fornecer informações sobre seus métodos, propriedades e eventos;
- Comunicação por eventos, um modelo de comunicação entre objetos alternativo a chamada de métodos;
- Customização, as propriedades de um *bean* podem ser configuradas de forma mais amigável;
- Persistência, as propriedades de um *bean* podem ser armazenadas de forma persistente.

Esse modelo de componente pode ser definido como um conjunto de classes na forma de pacotes Java, com o intuito de serem usados de uma maneira que permita isolar e encapsular um conjunto de funcionalidades. Ao contrário de outras tecnologias Java, os *beans* não precisam obrigatoriamente estender uma classe ou implementar uma interface (JEVEAUX, 2004).

As classes implementadas neste trabalho foram organizadas em um pacote chamado privperson. Tal pacote contêm classes Java e JavaBeans, e arquivos do tipo XML, os quais são utilizados para persistência das informações.

As classes Java compõem a "PrivPerson" sobre as funcionalidades que essa apresenta, como a criação do *crawler*.

O JavaBeans foi utilizado para inserção de informações, possibilitando abastecer a "PrivPerson" com novas ferramentas, cenários, camadas as quais esses contemplam, dentre outras informações de análise, recuperando as informações para a resposta ao usuário.

Existem algumas classes que são de grande relevância e merecem ser citadas, como as mostradas na Tabela 17.

**Tabela 17. Classes importantes na implementação**

| Nome da Classe | Descrição |
|---|---|
| PrivPerson.java | Contêm o menu e algumas das informações disponíveis na tela inicial. |
| ConsultaTaxonomia.java | Contêm métodos auxiliares para consultar as taxonomias definidas e seus respectivos pesos. |
| XmlModelPersistence.java | Biblioteca de métodos para manipulação das informações armazenadas nos arquivos do tipo XML. |
| AvaliarFerramenta.java | Provê a avaliação das ferramentas de privacidade e personalização. |
| AvaliarSites.java | Provê a avaliação dos sites. |
| PrivPersonCrawler | Responsável pelo *download* das páginas, inspeção e avaliação dos sites. |
| AvaliarCenario.java | Provê a avaliação dos cenários de utilização. |
| Sobre.java | Fornece informações de utilização e desenvolvimento da "PrivPerson". |
| GerenciarFerramentas.java | Classe restrita ao admin, responsável pelo gerenciamento das ferramentas na "PrivPerson" |
| GerenciarCenarios.java | Classe restrita ao admin, responsável pelo gerenciamento dos cenários de utilização na "PrivPerson". |

Tais classes foram separadas de acordo com a função atribuída a cada uma, de forma que se pudesse ter um código mais fácil de ser entendido, decorrendo de uma implementação facilitada e reutilização de código, se fosse necessária a mesma funcionalidade para classes diferentes.

Um exemplo de reutilização de código foi feito na utilização da classe xmlModelPersistence.java para a avaliação das ferramentas e cenários de utilização, a qual é responsável por gerenciar a manipulação dos arquivos XMLs para armazenamento e recuperação das informações.

Na implementação da "PrivPerson" também utilizou-se as classes, ou componentes *beans*, como apresentados na Tabela 18.

**Tabela 18. Classes beans desenvolvidas**

| Nome das Classes Componentes | Descrição |
|---|---|
| CamadaPesoPersonBean.java | Utilizado para manipulação das camadas da taxonomia de personalização e respectivos pesos. |
| CamadaPesoPrivBean.java | Utilizado para manipulação das camadas da taxonomia de privacidade e respectivos pesos. |
| DicaBean.java | Cadastro e recuperação das dicas disponibilizadas na tela inicial. |
| FerramentaBean.java | Cadastro e recuperação das ferramentas de privacidade e personalização disponíveis pela "PrivPerson". |
| CenarioBean.java | Cadastro e recuperação dos cenários disponíveis pela "PrivPerson". |

Foram utilizados arquivos do tipo XML para a persistência de informações que são necessárias e relevantes à analise e aos resultados disponibilizados aos usuários. Esse pode ser considerado como um banco de dados, no entanto, no formato de um texto de conteúdo variável, ao qual se aplica uma marcação que segue regras gramaticais bem definidas.

Foram utilizados arquivos XMLs durante a implementação da "PrivPerson" para facilitar a atribuição e recuperação de informações, através dos métodos de acesso do tipo *get* e *set*.

Para atribuição e persistência das taxonomias de privacidade e personalização e dos pesos atribuídos às camadas, utilizou-se também um arquivo XML, separado para cada taxonomia, de modo que se houver mudanças não seja preciso modificar o código fonte da "PrivPerson", já que tais informações não estão vinculadas ao código.

Uma funcionalidade importante apresentada pela "PrivPerson" é a capacidade de avaliação se o site utiliza *cookies* de forma automatizada. Isso pode ser feito através da análise do cabeçalho da página, podendo esse método ser visualizado na Figura 46.

```
public void readCookies(URL pageUrl, boolean printCookies,  boolean reset) throws IOException{
  URLConnection urlConn = pageUrl.openConnection();
   if (reset)
    theCookies.clear();
   int i=1;   String hdrKey;   String hdrString;   String aCookie;
   while ((hdrKey = urlConn.getHeaderFieldKey(i)) != null) {
    if (hdrKey.equals("Set-Cookie")) {
      hdrString = urlConn.getHeaderField(i);
      StringTokenizer st = new StringTokenizer(hdrString,",");
      while (st.hasMoreTokens()) {
        String s = st.nextToken();
        aCookie = s.substring(0, s.indexOf(";"));
        int j = aCookie.indexOf("=");
        if (j != -1) {
         if (!theCookies.containsKey(aCookie.substring(0, j))){
           theCookies.put(aCookie.substring(0, j),aCookie.substring(j + 1));
           addMatch(pageUrl.toString() + " Nome do Cookie: " + aCookie.substring(0, j) );
           if (printCookies){
            System.out.println("Lendo Chave: " + aCookie.substring(0, j));
            System.out.println("      Valor: " + aCookie.substring(j + 1));
           }
         }
        }
      }
    }
    i++;
   }
   urlConn = null;
}
```

**Figura 46. Método de análise da utilização de cookies**

Para acesso às páginas web de maneira automatizada, utilizou-se um *crawler*, o qual é constituído de um módulo da ferramenta chamado *"PrivPersonCrawler"*.

Esse é um robô de pesquisa web que faz análises nos sites disponíveis na Internet, verificando o IP das URL's informadas para análise e retornando a resposta da análise, caso a URL informada pelo usuário seja válida.

Para cada *link* encontrado nas páginas do site é armazenado seu endereço em uma tabela *hash*, possibilitando dessa forma que se mantenha um histórico de quais *links* ainda devem ser analisados. Dessa forma, todos os *links* referenciados pelo site, poderão ser avaliados, de forma a tornar a análise do *crawler* significativa e correta, evitando ainda análise repetida dos *links*.

Na avaliação manual seria preciso verificar todas as URLs referenciadas pelo site e ainda assim, comprovar, verificando na máquina do usuário se houve o armazenamento de *cookies*. Já na avaliação feita pela "PrivPerson" isso é feito de forma automatizada, analisando o cabeçalho do protocolo HTTP de cada página, de maneira rápida, tranqüila, segura e eficaz.

A "PrivPerson" ainda verifica algumas funcionalidades que podem ou não ser oferecidas pelos sites, o que os usuários leigos podem não saber como verificar e em muitas vezes nem conhecem, como por exemplo se o site apresenta ambiente seguro HTTPs.

Dessa forma a “PrivPerson” tenta reter o conhecimento de um especialista durante a análise dos mecanismos de modo a instruir os usuários em sua utilização.

A “PrivPerson” ainda faz tratamento de acentos nas palavras, tanto nas informações submetidas pelos usuários que a utilizam quanto nas informações recuperadas através dos sites, sendo essas informações importante à análise. Também são tratados os caracteres especiais do HTML aos quais os acentos são atribuídos às palavras, de forma que a pesquisa do *crawler* não seja prejudicada.

A “PrivPerson” foi desenvolvida com controle de possíveis erros dos usuários. Caso o usuário execute alguma ação incorreta, a “PrivPerson” emitirá uma exceção, sendo essa uma notificação do que ocorreu, em forma de janelas de aviso, instruindo o usuário no que deva ser feito.

A Tabela 19 mostra algumas exceções implementadas na “PrivPerson” para evitar erros que podem ser ocasionados pela não experiência dos usuários com a mesma.

**Tabela 19. Exceções implementadas**

| Nome da Exceção | Descrição |
|---|---|
| IOException | Exceção de entrada e saída, no caso de problemas de gravação e leitura do arquivo de *log* e na requisição HTTP. |
| startUrlException | Utilizada para verificar se a URL inicial, a qual o *cralwer* vai analisar, é diferente de vazio e válida. |
| maxUrlException | Se um número máximo inválido de URLs for digitado. |
| logFileException | Verifica se o caminho e nome do arquivo de *log* é válido. |

As exceções também foram utilizadas para garantir que o administrador não erre na utilização da “PrivPerson”. Por exemplo, se ocorrer erro na gravação das ferramentas, camadas, cenários de utilização, uma exceção é lançada sob a forma de mensagem de erro informando o ocorrido.

Como a “PrivPerson” foi implementada utilizando uma linguagem multi-plataforma, essa pode ser utilizada independente de ambiente operacional, possibilitando dessa forma, sua utilização por diferentes perfis de usuários.

## C.3. Ambiente para Desenvolvimento

Tendo decidido pela utilização da linguagem Java para a implementação da “PrivPerson” foi tomado como ambiente uma API que oferece suporte ao Java e disponibiliza recursos visuais, de forma a facilitar a criação de interfaces amigáveis.

Para isso utilizou-se a API NetBeans IDE 5.5 como ferramenta de desenvolvimento, o compilador Java e o ambiente de execução presente no pacote de desenvolvimento *Java Development Kit* (JDK) versão 1.5.0.09.

NetBeans IDE é uma ferramenta robusta e livre, escrita em Java, a qual permite desenvolvimento para diferentes aplicações. Foi desenvolvida pela Sun e pode ser encontrada gratuitamente através do endereço *http://www.netbeans.com*.

O JDK contém uma série de ferramentas para desenvolvimento em Java, mais ferramentas de compilação e depuradores (*debuggers*) necessários para desenvolvimento, podendo ser encontradas todas as versões e pacotes do JDK em *http://java.sun.com/jdk*. Existem outras características, apresentadas pela NetBeans IDE 5.5, que foram consideradas relevantes para a sua escolha como API de desenvolvimento, como:

- Suporte compreensivo para construção de módulos de *plug-in* IDE e ricas aplicações clientes na plataforma NetBeans;
- Construções GUI simplificadas e intuitivas usando o Project Matisse, o qual permite maior facilidade para desenvolvimento da interface pois disponibiliza um ambiente gráfico;
- Um extenso número de ações já disponíveis;
- Vários realces no ambiente de edição, incluindo faixa de erro e sugestões de código Java, facilitando a implementação e depuração do código;
- Conclusão de código mais rápida;
- Possui apoio de CVS, sendo esse um mecanismo de configuração de software, permitindo que se tenha um sistema de controle de versão;
- Depuração aumentada que é integrado no Editor de Fonte (*Source Editor*);

O ambiente operacional utilizado para a codificação, compilação e testes da "PrivPerson" foi composto por um microcomputador com processador AMD-Athlon(tm), 1.79 Ghz, com 512 MB de RAM, rodando em um ambiente computacional *Microsoft Windows XP Professional* - Versão 2002, *Service Pack* 2.

... uma nova caminhada se inicia aqui.

Luanna Lopes Lobato