

LEONARDO CAMARGOS DE ALMEIDA

**DESENVOLVIMENTO DE UM GERENCIADOR
DE ACESSOS A WEB SITES**

Monografia de graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências da disciplina Projeto Orientado para a obtenção do título de Bacharel em Ciência da Computação.

Orientador

Prof. Antônio Maria Pereira de Resende

LAVRAS
MINAS GERAIS - BRASIL
2001

LEONARDO CAMARGOS DE ALMEIDA

**DESENVOLVIMENTO DE UM GERENCIADOR
DE ACESSOS A WEB SITES**

Monografia de graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências da disciplina Projeto Orientado para a obtenção do título de Bacharel em Ciência da Computação.

APROVADA em 27 de junho de 2001.

Prof. André Luiz Zambalde

Prof. Olinda Nogueira Paes Cardoso

Prof. Antônio Maria Pereira de Resende
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	iv
1 INTRODUÇÃO	1
1.1 Internet	1
1.2 As empresas na Internet	2
1.3 A relação entre as empresas e seus clientes	3
1.4 A relação entre as empresas e seus clientes na Internet	4
1.5 Objetivo do trabalho	7
2 REVISÃO BIBLIOGRÁFICA	8
2.1 Personalização de web sites	8
2.2 Sites adaptativos	9
2.2.1 Definição dos Objetivos	10
2.2.2 Observação	11
2.2.3 Transformação	13
2.2.4 Aplicação	15
2.2.5 Exemplo	17
2.3 A coleta de informações sobre usuários	18
2.3.1 Análise de log	19
2.3.2 Serviços de registro de acesso	20
2.4 Tecnologias para desenvolvimento de registradores de acesso	22
2.4.1 Servlets	22
2.4.2 JSP	24
2.4.3 JDBC e ODBC	25
2.4.4 Web Server	25
3 MODELAGEM E IMPLEMENTAÇÃO DO GERENCIADOR DE ACESSO	27
3.1 Considerações iniciais	27
3.2 Informações sobre o sistema gerado	28
3.3 Diagramas de caso de uso	29
3.4 Diagrama de classes	29

4	VALIDAÇÃO DO GERENCIADOR DE ACESSOS	31
4.1	Instalação e configuração	31
4.2	Funcionamento e recursos	32
5	CONCLUSÕES E TRABALHOS FUTUROS	35
6	REFERÊNCIAS BIBLIOGRÁFICAS	36
A	Amostras de Telas do Sistema	39

LISTA DE FIGURAS

Figura 1 - Processo de personalização de um web site	10
Figura 2 - Diagrama de casos de uso do sistema RAJ	29
Figura 3 - Diagrama de classes do projeto	30
Figura 4 - Amostra da tela de exibição do número de acesso das páginas registradas de um web site	39
Figura 5 - Amostra de parte da tela de exibição de sistemas operacionais utilizados pelos usuários de um web site	39
Figura 6 - Amostra de parte da tela de exibição de browsers utilizados pelos usuários de um web site	40
Figura 7 - Amostra de parte da tela de exibição dos hosts dos usuários .	40
Figura 8 - Amostra de parte da tela de configuração de informações do web site	41
Figura 9 - Amostra de parte da tela de exibição de páginas monitoradas pelo serviço de registro de acesso	41
Figura 10 - Amostra de parte da tela de cadastro de página a ser monitorada pelo serviço de registro de acessos	41

LISTA DE TABELAS

Tabela 1 - Preços do serviço IDStat	21
Tabela 2 - Preços do serviço Jinko	22
Tabela 3 - Preços do serviço WatchWise	22

CAPÍTULO 1

INTRODUÇÃO

1.1 Internet

Inicialmente, a Internet era usada somente por um pequeno grupo de pessoas nos meios acadêmicos para troca de informações científicas. Entretanto, hoje é possível encontrar quase todo tipo de informação dentro da Internet: "A Internet é agora um dos maiores repositórios mundiais de informação" [CHA95]. Esta frase define bem o potencial da rede mundial de computadores no mundo globalizado em que se vive atualmente. Pessoas das diferentes nacionalidades, falando as mais diversas línguas, com interesses e conhecimentos variados podem trocar e compartilhar informações, colocando suas informações disponíveis publicamente e, ao mesmo tempo, procurando por conhecimento e experiências que outros adquiriram e vivenciaram.

Na sua forma mais conhecida e utilizada, a *World Wide Web* (que representa o conjunto de documentos multimídia da Internet) pode ser vista como um "grande e autônomo sistema de informações distribuído e heterogêneo, utilizado como fonte de informações nas mais diversas áreas do conhecimento, seja por motivos profissionais ou simplesmente por lazer" [SSU96].

Desde o seu início, além da expansão do conteúdo da Web, também as tecnologias utilizadas foram evoluindo de forma contínua e rápida, tornando a navegação mais interessante. Além disso, o número de usuários com acesso a grande rede mundial de computadores cresceu significativamente. Só para se ter idéia da expansão no Brasil, o "número estimado de usuários na Internet, para o ano de 1999, variou entre 4 e 7 milhões dependendo da fonte. Em números absolutos, o Brasil ocupa lugar significativo em termos de usuários - variando entre o 12º e 14º lugar no *ranking* mundial, a depender do critério, ao passo que,

em números relativos, a quantidade de usuários em relação ao total da população coloca o país em 4º lugar na América Latina” [TAK00].

1.2 As empresas na Internet

O crescimento acelerado da Web e a possibilidade de disseminar muita informação pelo mundo motivaram grandes empresas a apostarem no potencial comercial da Internet. Vencendo pouco a pouco a resistência dos usuários, as empresas foram inicialmente oferecendo sites de divulgação de informações e utilidade pública. Com a evolução, desenvolveram-se outras atividades na Web, tais como, *e-commerce*, *e-business*, *e-gov*, etc.

Um sistema de comércio eletrônico pode ser definido pela "habilidade de se vender e comprar bens eletronicamente, através de uma rede" [TYG97]. Para permitir que potenciais compradores confiassem o suficiente na segurança da informação, por exemplo, para utilizarem o número de seus cartões de crédito nas transações virtuais, muitas empresas investiram em técnicas de criptografia.

Com o passar do tempo e o aumento do número de empresas as quais disponibilizam seus produtos e serviços na Web, criou-se a mentalidade de que as empresas fora da Internet iriam desaparecer em pouco tempo, esmagadas pela concorrência eletrônica. Este pensamento, que agora se revela errôneo, impunha uma única alternativa para as corporações: invadir a Web. Esta corrida pela conquista da Web, provocou uma explosão de empresas na Internet. Neste primeiro momento, a Internet figurava subsidiariamente no oferecimento de produtos e serviços já existentes em lojas físicas. Assim, surgiu o conceito de empresa virtual, como é conhecido hoje, uma empresa que existe somente no espaço virtual, não tendo instalações físicas.

1.3 A relação entre as empresas e seus clientes

Há muito tempo, as empresas compreendem que o bom atendimento e a satisfação dos clientes são prioridades na realização das atividades da empresa. O bom atendimento e a satisfação do cliente compõe a característica qualidade [FEI91]. Surge daí a necessidade de informações a respeito dos consumidores para melhor atender e satisfazer os clientes. Conhecendo as preferências, perfil e padrão de consumo dos clientes, os administradores podem traçar planos e estratégias para melhorar o empreendimento, atrair e conquistar mais consumidores e, conseqüentemente, aumentar o lucro da empresa.

Na busca dessas informações, pesquisas podem ser realizadas a fim de coletar dados dos clientes. Por exemplo, questionários respondidos por consumidores podem fornecer informações sobre assuntos específicos que intrigam os administradores. Além disso, a observação de sugestões e reclamações podem ser úteis para alterar determinados pontos onde hajam falhas na empresa.

Entretanto, determinadas informações não fornecidas diretamente pelos clientes poderiam ser bastante proveitosas para a empresa. Por exemplo, em [AIS93] é discutido o caso da rede de hipermercados Wal-Mart que, ao fazer um estudo em seu banco de dados, descobriu que o perfil do consumidor de cervejas era semelhante ao perfil do comprador de fraldas. Eram homens casados, entre 25 e 30 anos, que compravam fraldas e/ou cervejas à tarde no caminho do trabalho para a casa. Com base nisso, o Wal-Mart optou por uma reestruturação das prateleiras, colocando as fraldas ao lado das cervejas. Como resultado, o consumo cresceu 30% às sextas-feiras devido a redefinição da organização, baseada na conexão de hipóteses desenvolvidas pela mineração de dados em seu banco de dados.

1.4 A relação entre as empresas e seus clientes na Internet

Da mesma forma que os negócios no mundo real, as empresas que desejam prosperar na Web necessitam de informações sobre os usuários de seus produtos e serviços. Assim, os web sites em conformidade com as expectativas e exigências de seus visitantes têm muito mais chances de atingir seus objetivos, sejam eles comerciais ou não.

A fim de atingir seus objetivos, deve-se estruturar um web site da melhor forma para agradar e conquistar os usuários. A seguir, citamos alguns exemplos, fazendo um paralelo entre organização física de uma empresa.

Exemplo 1 - Suponha uma grande loja de departamento com vários andares onde não há indicações sobre a localização de suas seções. Um novo cliente, que ainda não conhece a disposição das seções na loja, percorre todo o estabelecimento até encontrar o produto que realmente precisa. Entretanto, pela dificuldade de acesso à mercadoria, este cliente pode desistir de procurá-la, dirigindo-se a outra loja para adquiri-la. Da mesma forma, um web site de uma loja de departamentos que não contém um mapa ou indicação de seu conteúdo, pode ser confuso para um cliente que procura alguma informação. Obrigar o usuário a percorrer diversas páginas, pode levá-lo a desistir de procurar o que deseja ou até mesmo a buscar em um site similar.

Exemplo 2 - A disposição de artigos em lojas de CDs pode levar em consideração o gênero musical ou a semelhança entre os artistas. Assim, quando o cliente procurar um determinado item, ele pode comprar, não somente o CD que o levou a loja, mas também algum outro similar que também o agrade. Pensando em termos de um catálogo de CDs na Internet, pode-se disponibilizar nas páginas informações de CDs, agrupadas por artista, estilo musical, etc. O visitante ao examinar a página pode se interessar e comprar outro item, mesmo não sendo esse o seu propósito inicial.

Exemplo 3 - Uma estratégia muito utilizada, principalmente em supermercados, é a organização dos produtos de forma que o cliente deva percorrer seções de supérfluos para chegar aos produtos de necessidade básica. Caminhando entre as prateleiras de mercadorias, o cliente pode ser tentado a adquirir produtos "desnecessários" em sua lista básica de compras. Normalmente, observa-se que os produtos da cesta básica são colocados bem ao fundo do estabelecimento, obrigando o consumidor a passar por seções atrativas que o instigam a adquirir doces, bebidas, petiscos, etc. A mesma estratégia pode ser utilizada em um site onde se conhece qual o principal motivo da visita do usuário. Fazendo com que o cliente visualize anúncios e páginas de outros produtos antes de chegar na informação realmente desejada, pode-se conseguir o seu interesse na aquisição de outras mercadorias.

Nos exemplos apresentados acima, a maior parte do comportamento esperado do cliente pode ser deduzido pelo fornecedor dos produtos e serviços no momento da estruturação do web site. Entretanto, é necessário um meio de se medir e analisar a eficiência dessa estruturação. Por outro lado, há uma série de questões relativas ao comportamento navegacional do usuário que não podem ser deduzidas e que se fossem conhecidas poderiam ser bastante úteis para guiar a manutenção de recursos, prover serviços personalizados e auxiliar em estratégias de *marketing*. A seguir, apresenta-se alguns casos da utilidade de informações sobre o comportamento de usuários dentro de um web site específico.

Exemplo 4 - Ao conhecer os períodos de maior tráfego em um site de comércio eletrônico onde ocorre um grande volume de transações, pode-se distribuir as requisições a fim de aliviar a carga de processamento dos servidores e diminuir falhas no serviço e tempo de resposta. No caso especial de comércio eletrônico, essas medidas tornam-se fundamentais pois já foi identificado que o uso de comércio eletrônico foi desmotivado devido a congestionamentos na rede

e baixo desempenho. Muitos sites de comércio eletrônico têm enfrentado problemas e perdas financeiras quando os usuários são impedidos de negociar em um tempo adequado, ou seja, são impedidos de fechar suas compras ou mesmo de adicionar um item ao seu carrinho de compras. Esses problemas afastam os usuários do site e, conseqüentemente, diminuem as possibilidades de lucro do mesmo.

Exemplo 5 - Gerar relatórios estatísticos usando as informações sobre os acessos e apresentar os resultados, tanto a investidores quanto ao pessoal de desenvolvimento da empresa, a fim de exibir um reflexo do progresso do produto ou serviço. Com isso, pode-se gerar estímulos e maior interesse.

Exemplo 6 - Se, ao analisar os registros de visitantes de um web site, for descoberto que grande parcela da visitação é de um país estrangeiro, isto em contra partida, deverá influenciar ao acréscimo de suporte a outra língua, adaptando o web site para estar compatível com os possíveis clientes e aumentando as vendas.

Portanto, pode-se perceber que informações sobre os visitantes tornam-se necessárias aos administradores e responsáveis no processo de tomada de decisão a respeito de um empreendimento virtual. De posse dessas informações, pode-se traçar diretrizes a fim de melhorar, expandir ou direcionar o conteúdo do web site. Desta forma e cada vez mais, as exigências e expectativas dos usuários poderão ser satisfeitas.

1.5 Objetivo do trabalho

O presente trabalho pretende discutir um modelo de processo de personalização de web sites utilizando informações sobre acessos de usuários para melhorar a estrutura das páginas e a organização do conteúdo.

Uma primeira versão de um serviço de gerenciamento de registro de acessos será desenvolvido a fim de coletar e gerar estatísticas a respeito dos usuários de um web site. O serviço ficará disponível na Internet e poderá ser usado pelos administradores para conhecer informações sobre os visitantes de seu web site, como por exemplo, as páginas mais e menos acessadas, os navegadores e sistemas operacionais utilizados, datas de acesso, *hosts*, etc.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

2.1 Personalização de web sites

Projetar um web site e organizar o seu conteúdo não é uma tarefa trivial. Normalmente um web site condensa diversas informações distribuídas entre as páginas, imagens e *hiperlinks*. Mesmo que a audiência do site não seja grande, geralmente, o público é diversificado. Cada pessoa que acessa um site pode ter um objetivo diferente, pode estar procurando informações diferentes. É bem provável que a maioria dos usuários da Internet tenha visitado sites que poderiam ter a informação, o produto ou o serviço que ele estava procurando, mas não conseguiu encontrá-los e se dirigiu para um outro site.

Ao mesmo tempo que projetar um web site é uma tarefa complicada, que depende de vários fatores, deve-se conhecer o público que ele tem. Por exemplo, precisa-se de um recurso que descreva as atividades dos usuários que estão navegando pelo site. Sabe-se que qualquer servidor web mantém o log de todas as requisições recebidas, com a data, hora e o endereço IP de onde as requisições foram enviadas. Dessa forma, os usuários deixam "rastros" enquanto navegam por um site. E, aplicando técnicas apropriadas para extração destas informações, este site pode melhorar a sua estrutura.

Uma forma de aproveitar as informações deixadas pelos usuários enquanto navegam pelo site é inferir o perfil, padrão de comportamento e interesses dos usuários e adaptar a estrutura do site de acordo com esse aprendizado. Mineração de dados, redes neurais e algoritmos de recuperação de informação são exemplos de técnicas utilizadas para extrair informações dos dados dos usuários. Algumas possíveis consequências da aplicação dessas técnicas são: aumento da média de páginas visitadas por sessão, os usuários

podem demorar mais tempo visitando o site, pode haver um aumento da taxa de retenção (número de visitantes que retornam ao site) e da taxa de conversão (visitas que se tornam compras) [GRE00].

2.2 Sites adaptativos

Neste ponto, pode-se definir sites adaptativos [PET97]: “web sites que utilizam as informações sobre o padrão de acesso dos seus usuários para melhorar a estrutura das suas páginas e a organização do seu conteúdo”. Define-se como personalização de web sites o processo de transformar um site em adaptativo. Sites adaptativos observam as atividades dos usuários, os seus erros e aprendem sobre os perfis de usuário, sobre os seus padrões de acesso e problemas com a organização do conteúdo de um site. A personalização de web sites é uma estratégia para aproveitar as informações deixadas pelo usuário com o objetivo de tornar o site mais próximo das necessidades do seu público.

A seguir, apresenta-se uma taxonomia discutida em [RRF99] que classifica as técnicas de personalização existentes e define todo o processo de personalizar um site. O processo de criação de um site adaptativo envolve quatro fases (Figura 1):

- definição dos objetivos: o administrador do site deve definir as metas a serem alcançadas com a personalização;
- observação: a coleta dos dados de acesso ao site;
- transformação: gera as regras de adaptação a partir dos dados coletados na fase anterior;
- aplicação: a utilização das regras na sua estrutura.

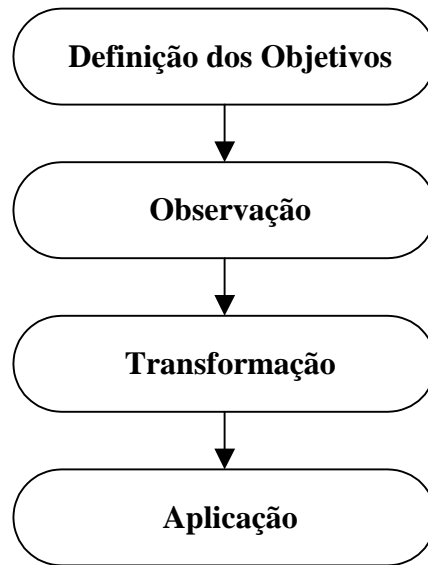


Figura 1 - Processo de personalização de um web site

2.2.1 Definição dos Objetivos

O primeiro passo para aplicar estratégias de personalização em um site é definir os objetivos a serem alcançados com a aplicação dessas estratégias. Somente com esses objetivos definidos é que as estratégias poderão ser projetadas. O administrador de um site pode ter diversos motivos para decidir aplicar estratégias de personalização em um web site. Pode-se classificar esses motivos da seguinte forma:

I) Funcionais

Neste caso, o administrador do site quer facilitar a visita do usuário. A estrutura do site deve ser rearranjada de forma que o usuário encontre o que estiver procurando ou que acesse seções que o interessam e com facilidade.

II) Comerciais

Neste caso, se o objetivo é fomentar a venda de um determinado produto. O administrador espera que depois da personalização aumente o número de vezes que uma ação ocorra, por exemplo, pagar o produto “A”.

2.2.2 Observação

Considerando um web site, a necessidade de torná-lo adaptativo e os objetivos a serem alcançados com a personalização, o próximo passo do processo deve ser a coleta de dados.

Na fase de observação, o objetivo é coletar os dados relacionados aos usuários do site e, opcionalmente, dados relacionados ao negócio do site. Deve-se ressaltar que não é possível realizar uma análise da eficiência da estrutura do site, com o fim de alterar esta estrutura, se não existem dados sobre interação dos usuários com o mesmo. Neste ponto, pode-se considerar as seguintes perspectivas:

I) Identificação dos usuários

Define se os dados coletados serão utilizados para identificar o usuário.

- **Usuário é identificado:** nesse caso, é possível guardar as informações de cada usuário, e a cada vez que ele retornar ao site, estas informações podem ser aproveitadas. É possível analisar o comportamento dos usuários ao longo de um período de tempo e perceber quais partes do site estão sendo descobertas e em qual ordem [FKN96].

- **Usuário não é identificado:** não é necessário para tornar um site adaptativo que os seus usuários sejam identificados. Quando a forma de aquisição dos dados é através do log do servidor web, pode ser que apenas a identificação das sessões seja suficiente. Essa situação pode ser comparada às

análises realizadas com os dados de compra de um supermercado, onde a identidade do comprador não é considerada quando se aplica as técnicas de mineração de dados [FSS99].

II) Fonte de dados

Trata da natureza dos dados coletados. Especifica quais tipos de dados serão utilizados na próxima fase, a de transformação.

- Dados de acesso dos usuários: para descobrir detalhes do comportamento dos usuários do site, é suficiente conhecer as páginas acessadas em cada sessão e a ordem em que elas foram acessadas. As preferências dos usuários podem ser inferidas através das páginas do site que ele acessou. Por exemplo [RRF99], dado que em várias sessões, uma dada página é acessada depois que um determinado conjunto de páginas é acessado, toda vez que este conjunto de páginas for acessado, o site pode mostrar um *link* para a página.

- Dados pessoais dos usuários: se informações pessoais dos usuários do site estiverem disponíveis - como profissão, cidade/região/país onde mora, estado civil - outros tipos de análises podem ser realizadas. O site pode descobrir com estas análises que o perfil dos usuários do site é, por exemplo, homem mineiro, advogado e casado. Com essa informação, pode destacar os *links* para as páginas relacionadas a Direito.

- Dados do mercado: utilizar dados de um *data warehouse*, como por exemplo os dados das vendas realizadas pela empresa que não foram feitas via Internet [GHL00]. Informações de quais produtos são vendidos juntos podem ser obtidas neste *data warehouse* e aproveitadas nas vendas via Internet.

III) Forma de aquisição dos dados de acesso

Especifica a forma como os dados são coletados, definindo se este processo é ou não transparente ao usuário.

- Não transparente: o usuário, voluntariamente, pode fornecer dados para o site. Por exemplo, o site pode fazer perguntas ao usuário para saber qual o seu objetivo naquela sessão e saber se ele foi ou não alcançado. Assim, pode criar um histórico de acessos, baseado nos sucessos e insucessos dos usuários [JFM97].

- Transparente: o site pode coletar as informações de acesso do usuário sem que ele precise responder qualquer tipo de questionário. Num arquivo de log estão todas as requisições feitas ao servidor web com o endereço IP da máquina que fez as requisições e a data/horário do acesso. É possível identificar sessões e identificar qual a ordem de acesso das páginas de um site, durante uma sessão. Com estas informações, já é possível realizar alguns tipos de análises [FSS99].

2.2.3 Transformação

Depois da fase de observação, ou seja, depois que dados suficientes já foram coletados, precisamos descobrir padrões de acesso (ou de comportamento), os perfis dos usuários, tendências, para que a estrutura do site possa ser melhorada. Os dados coletados na fase de observação precisam ser trabalhados, relacionados e bem entendidos para que informações válidas sejam produzidas.

É nesse contexto que entra em cena a transformação, cujo objetivo é obter informações a partir de dados. Nesta fase, várias decisões caracterizam a estratégia:

1) Técnicas de análise

Identifica quais técnicas de análise serão utilizadas para explorar os dados.

- Padrões Informativos: quando não se tem um problema específico, uma forma de se obter informações de uma grande massa de dados é utilizando técnicas para descobrir padrões informativos. Os algoritmos de agrupamento são bons exemplos para este tipo de padrão. Eles identificam características comuns a um conjunto de entradas de forma a particionar a base de dados de acordo com os seus atributos. Uma possível aplicação nesse caso, seria descobrir os perfis de usuários que acessam um site, com base nas páginas visitadas e nas palavras pesquisadas [MDU99], ou mesmo com as suas características pessoais.

- Padrões de previsão: dada uma entrada de dados e os seus atributos, padrões de previsão têm o objetivo de fazer uma estimativa razoável de um atributo desconhecido com base em outros atributos conhecidos. Redes neurais e regressão linear vêm sendo amplamente utilizadas para determinação de padrões de previsão e também podem ser usados para gerar as regras de transformação do site.

II) Nível de automatização

Define se a fase de transformação estimula ou não que o administrador do site participe nas deduções sobre a personalização.

- Não automatizado: uma abordagem interessante é que o sistema estimule a interação do administrador do site em momentos em que as heurísticas aplicadas não conseguem decidir qual caminho seguir. Por exemplo, um sistema pode identificar que determinados grupos de páginas são acessadas simultaneamente, e sugere então ao administrador do site que seja criada uma página no site com *links* para essas páginas. Nesse caso, o administrador do site ficaria encarregado de descobrir a que assunto essas páginas estão relacionadas e dar um título para a nova página do site [PET99].

- Automatizado: se a técnica de transformação dos dados não depender da interferência humana, o processo exige uma menor interação do administrador do site que, em alguns casos, pode ser a melhor opção [PET98].

2.2.4 Aplicação

Durante a fase de transformação, obteve-se conhecimento sobre os padrões de acesso ao site e o que fazer para melhorá-lo. Na fase de aplicação, deve-se aplicar este conhecimento adquirido na estrutura do site. Neste ponto, pode-se considerar as seguintes perspectivas:

I) Escopo das aplicações

Define qual o alcance das transformações.

- Global: nesse caso, o site aprende com os dados dos usuários e faz alterações permanentes na estrutura do site. Quando a fase de transformação, por exemplo, sugere a criação de uma nova página, a aplicação da técnica então altera definitivamente a estrutura do site [PET98].

- Local: o site é dinâmico e se adapta à sessão/usuário corrente. Quando a técnica utilizada apenas inclui *links* não permanentes na página que está sendo acessada, esta alteração não é definitiva. A cada sessão, a estrutura do site pode sofrer transformações diferentes, que serão desfeitas assim que aquela sessão terminar [RRF99].

II) Automatização

Define se as mudanças feitas no site necessitam da autorização do administrador do site.

- As transformações são aplicadas automaticamente: nesta abordagem, o site pode interagir diretamente com o usuário, alterando as páginas pelas quais está navegando, sem que haja a permissão de um administrador do site [RRF99].

- As transformações são apenas sugeridas: o site pode apenas sugerir mudanças na estrutura para o administrador do site e ele as aplica se achar conveniente. Esta abordagem geralmente leva a transformações definitivas na estrutura do site [PET98].

III) Recurso de formatação

Identifica quais recursos de formatação das páginas são utilizados na aplicação das mudanças.

- Proprietário: já foram desenvolvidos alguns tipos especiais de linguagens de marcação, destinado a sites adaptativos. Uma página pode ser construída com *tags* especiais no HTML e as adaptações são feitas nas regiões delimitadas por essas *tags*. Por exemplo, essa *tag* pode ter argumentos que indicam que um *link* deve ser destacado, caso a página seja visualizada por um determinado usuário, num determinado contexto [PET97].

- Público: o HTML ou XML padrão são utilizados e os recursos de adaptação são feitos utilizando as *tags* já existentes [BBA99].

IV) Amplitude das transformações

Identifica quais regiões do site serão afetadas durante a fase de aplicação.

- As regiões "adaptáveis" são fixas, pré-determinadas: o site pode, por exemplo, ter um *frame* superior onde sempre aparecem *links* relacionados com a página corrente ou os *links* que as pessoas seguem depois que acessam a página corrente. Esta é uma forma simples de aplicar o conhecimento adquirido na fase de transformação [JFM97].

- As adaptações podem ocorrer em qualquer lugar do site: nesse caso, nenhuma região do site é destinada somente a adaptações. Essas podem ocorrer em qualquer parte da estrutura do site [CHA99].

2.2.5 Exemplo

Conforme ilustrado na figura 1, o processo de personalização envolve três fases: definição dos objetivos, observação, transformação e aplicação. Depois de executada a fase de transformação, a estratégia de personalização está definida. Portanto, as regras para adaptar o site de acordo com a sessão do usuário já estão prontas, mas ainda não estão implementadas no site. É neste momento que uma ferramenta para personalização deve ser utilizada.

Para exemplificar, considere que uma livraria virtual tenha percorrido todo o processo de personalização até a fase de transformação e que se agora utilize a ferramenta para estimar como ficará a visitaç o do site depois da personalizaç o.

Segundo a taxonomia apresentada, a fase de definiç o de objetivos pode ser classificada em funcional e comercial. Com o objetivo funcional, a meta   melhorar o caminharmento do usu rio no site, levando-o onde ele quer ir. Assim, o objetivo   criar *links* entre p ginas para diminuir o n mero de p ginas visualizadas pelo usu rio para ele encontrar o que est  procurando. J  o objetivo comercial escolhido foi aumentar a venda de livros.

Durante a fase de observaç o, foram coletados tanto dados sobre o padr o de acesso dos usu rios, como dados de compra de usu rios no site. No primeiro caso, n o foi considerado um identificador de usu rios, j  que muitos usu rios podem acessar o site sem se identificar. Na coleta dos dados de compras, foi mantido o identificador do usu rio que fez a compra.

A fase de transformação gerou um conjunto de regras para cada objetivo. Para o objetivo funcional, foram encontradas regras do tipo em x% das vezes que uma pessoa seleciona o livro “A”, ela também seleciona o livro “B”. Para o objetivo comercial, a regra encontrada dizia em y% das vezes que uma pessoa compra o livro “C”, numa outra visita ela acaba comprando o livro “D”.

De acordo com o processo descrito, foram definidas duas estratégias de personalização para a livraria virtual:

- a) Ao selecionar um livro, verifique se aquele livro leva à seleção de outro livro. Se levar, crie um *link* para a página de selecionar o livro.
- b) Quando o usuário for identificado no site, verifique se ele já fez alguma compra e se há algum livro para recomendar. Se houver, crie um *link* para a página de selecionar o livro.

2.3 A coleta de informações sobre usuários

Para personalizar um site, torna-se necessário fazer uso de algum processo de coleta de dados para obter informações a respeito dos usuários. Para descobrir as preferências, perfil e padrão de consumo dos clientes em potencial de um produto ou serviço, pode-se realizar pesquisas de opinião, formulando-se perguntas para um questionário que deve ser aplicado a uma amostra da população.

No caso de web sites, é bastante evidente que o melhor instante e local para coletar dados de usuários, é realizar a coleta exatamente no momento de acesso do conteúdo do site; desta forma, têm-se a certeza da coleta de dados de uma parcela representativa das pessoas que são realmente clientes do produto ou serviço.

Entretanto, pode-se ter problemas com o uso de questionários interativos para pesquisas na Internet, uma vez que, por falta de tempo ou desinteresse em

contribuir, algumas pessoas podem optar por não respondê-los ou propositalmente fornecer respostas erradas. Assim, os resultados da pesquisa não corresponderão a realidade. Portanto, muitas vezes (quando for possível), a melhor solução é tornar a coleta de dados automática e transparente ao usuário, de forma que os dados sejam capturados sem sua intervenção.

2.3.1 Análise de log

Uma abordagem para o descobrimento de informações sobre os usuários de um web site seria a análise dos arquivos de *log* do servidor Web. Estes arquivos contêm um histórico completo dos arquivos acessados pelos clientes. O *log* é um mecanismo do servidor Web que armazena registros para cada requisição vinda dos visitantes. A maioria dos *logs* de acesso WWW segue o formato CLF (*Common Log Format*) especificado como parte do protocolo HTTP. Cada entrada em um arquivo de *log* tem três campos importantes: o endereço IP de onde a requisição se originou, a descrição da requisição e uma indicação do momento esta ocorreu.

Como resultado das interações entre visitantes e um web site [TMU00], um arquivo de *log* contém muita informação sobre os comportamento dos usuários dentro do site, que se totalmente explorada, pode melhorar os serviços para o cliente e o desempenho do site .

Há muitas ferramentas de análise de *log* disponíveis na Internet, tais como, ilyx Edge [ILU99], Marketwave [MAR98], Webalizer [BAR01] e WebTrends [WEB99]. Estas ferramentas geram dados estatísticos a partir dos dados brutos armazenados nos arquivos de *log*, resultando em relatórios de totais de visitas às páginas, páginas mais acessadas, data e hora de utilização, etc. Estas informações são úteis em um cenário onde os dados a serem analisados são resultados de fatores quantitativos, tais como, contagem de acessos por domínio e quantificação de uso de um recurso. Entretanto, o desempenho destas

ferramentas cai rapidamente quando o tamanho do conjunto de dados aumenta [TMU00]. Além disso, grande parte das pessoas que não têm seus web sites hospedados em servidores dedicados, não têm permissão para acessar os arquivos de *log* do servidor web, uma vez que, estes arquivos também contém informações estratégicas sobre todos outros os web sites hospedados naquele mesmo servidor. Mas a principal desvantagem é a necessidade de se fazer a análise do *log* toda vez que as informações mais recentes sobre os usuários do web site forem requeridas. Um outro problema é que, normalmente, essas informações não ficam disponíveis *online* para serem consultadas em qualquer lugar onde haja um computador conectado à Internet.

2.3.2 Serviços de registro de acesso

Uma outra abordagem para a coleta de informações sobre os usuários do site é o uso de serviços de registro de acesso disponíveis na Internet. Estes serviços se apresentam em forma de web sites e registram o acesso dos usuários a sites previamente cadastrados. Alguns exemplos de tais serviços são HWCounter [HIG01], SuperStats [MYC01], IDStat [MAS01], Jinko [ROO01] e WatchWise [WAT01].

O uso deste tipo de serviço não exige nenhuma programação, nem é preciso instalar qualquer arquivo ou programa para ter o registro de acesso dos usuários sendo executado no site do cliente. Todas as funções de coleta e geração de estatísticas são realizadas remotamente pelo fornecedor do serviço, com isto é possível rastrear, compilar e tornar as informações disponíveis em tempo real, enquanto os visitantes estão acessando o site. Para ter o registro dos acessos sendo realizado no site, somente é preciso inserir nas páginas um pequeno código HTML o qual é fornecido quando se contrata o serviço. Este código, quando colocado num documento HTML, faz uma requisição ao servidor do fornecedor sempre que a página for carregada. Neste processo,

informações sobre o visitante são extraídas desta requisição e armazenadas em arquivos específicos pelo servidor.

Além de diversos tipos de relatórios sobre os acessos, alguns serviços também oferecem recursos adicionais, tais como: envio de relatórios [HIG01] por e-mail em diversos formatos; monitoramento 24h por dia [WAT01] que verifica se o web site está no ar e envia um alerta, via e-mail, caso contrário; etc.

A maioria destes serviços disponibiliza seus recursos gratuitamente por um curto período de avaliação que varia de 3 dias [ROO01] a 1 mês [MAS01]. Dessa forma, é possível avaliar a utilidade e as vantagens oferecidas e, terminado este período, pode-se decidir pela continuidade do serviço sendo, então, obrigatório o pagamento.

Geralmente, cada fornecedor deste tipo de serviço estabelece um preço que depende de algum critério escolhido, como, por exemplo:

- Número de acessos ao site, sendo o valor cobrado relativo a faixa de acessos obtidos [MAS01];
- Tipo do plano de serviço escolhido pelo cliente, sendo que cada plano apresenta recursos que podem não estar disponíveis em outros planos [ROO01];
- Forma de pagamento pretendida pelo usuário, sendo o pagamento mensal ou anual os mais comuns [HIG01] [WAT01];
- A natureza do site, geralmente classificados em comerciais ou sem fins lucrativos [WAT01].

Alguns exemplos de preços e critérios para cálculo de pagamento:

Tabela 1 - Preços do serviço IDStat [MAS01]

Número de Acessos obtidos	Custo anual (em US\$)
0 - 15.000	48,00
15.001 - 30.000	78,00
30.001 - 50.000	108,00

Tabela 2 - Preços do serviço Jinko [ROO01]

Tipo de conta	Custo anual (em US\$)
Conta Básica	24,00
Conta Detalhada	36,00
Conta Universal	84,00
Conta Avançada	144,00

Tabela 3 - Preços do serviço WatchWise [WAT01]

Categoria do Web site	Custo semi-anual (em US\$)	Custo anual (em US\$)
Sites não-comerciais (páginas pessoais e empresas sem fins lucrativos)	50,00	99,00
Sites comerciais (pequenos negócios e corporações)	150,00	299,00

2.4 Tecnologias para desenvolvimento de registradores de acesso

Para criar um serviço que registre acesso de usuários a web sites, é necessário utilizar algumas tecnologias para o desenvolvimento da aplicação. Algumas tecnologias possíveis são: CGI, ASP, JSP, PHP, servlets, entre outras.

A seguir, é apresentada uma breve discussão sobre as tecnologias utilizadas na implementação do serviço discutido neste trabalho.

2.4.1 Servlets

Servlets são aplicativos escritos em Java que podem ser acoplados em diversos tipos de servidores Web para expandir as suas funcionalidades. Os *servlets* são a solução em Java para a programação CGI: são programas que rodam em servidores Web e permitem a geração dinâmica de páginas HTML,

além de poder incluir acesso a banco de dados ou comunicação com outros servlets. Ao contrário dos *applets*, não possuem uma interface gráfica, e podem ser escritos para trabalhar com diversos tipos de protocolo de comunicação, como SMTP, HTTP e IRC.

Os *servlets* Java são mais fáceis de usar, mais eficientes, poderosos e portáveis do que CGI's tradicionais ou alternativas similares [HAL00]. Com CGI, um novo processo é iniciado para cada requisição HTTP. Se o programa CGI realiza uma operação relativamente rápida, o desperdício de tempo ("*overhead*") para iniciar o programa pode determinar o tempo de execução. Com servlets, a máquina virtual Java (*Java Virtual Machine*), após iniciada, permanece na memória e cada requisição é tratada por um processo ("*thread*") Java, muito mais leve que uma operação de sistema. Além disso, servlets tem muito mais alternativas para otimizações tais como armazenamento de computações anteriores ("*caching*") e conservação de conexões a banco de dados.

Servlets permitem facilmente diversos recursos que seriam difíceis ou impossíveis de serem feitos com CGI. Tem uma infra-estrutura abrangente para decodificar e analisar dados de formulários HTML, ler e ajustar cabeçalhos HTTP, lidar com *cookies*, rastrear sessões e muitas outras utilidades. Por exemplo, *servlets* podem interagir diretamente com o servidor Web, o que programas CGI não podem fazer. Isto simplifica operações que precisam procurar por imagens e outros dados armazenado em locais padrão.

Como os *servlets*, são escritos em Java e seguem uma API bem padronizada, são bastante portáveis. Assim, um servlet pode rodar sem alterações em qualquer servidor que suporte essa tecnologia.

2.4.2 JSP

Java Server Pages (JSP) é uma tecnologia que permite misturar HTML estático com HTML gerado dinamicamente. A vantagem do JSP consiste no fato da parte dinâmica ser escrita em Java, sendo esta uma linguagem poderosa e fácil de usar. Além de apresentar portabilidade para outros sistemas operacionais, sua execução não é dependente de servidores Microsoft como ocorre com a tecnologia ASP similar a esta.

Muitas páginas na Web geradas por CGI's tem grandes partes estáticas com partes dinâmicas limitadas a poucos e pequenos locais. A maioria da alternativas para CGI, incluindo servlets, exige que páginas inteiras sejam geradas por programas, embora a maior parte delas seja sempre a mesma. O JSP permite separar a parte dinâmica das páginas do código HTML estático.

Embora uma página em JSP seja muito parecida com um arquivo HTML comum, nos bastidores, a página JSP é convertida em um *servlet* onde o código HTML estático é impresso por uma fluxo de saída associado. Esta conversão é feita da primeira vez que a página é requisitada. Normalmente, os desenvolvedores fazem uma requisição a todas as páginas JSP, após a instalação, para terem certeza de que o primeiro usuário real não terá um atraso momentâneo decorrente do tempo necessário para que a página JSP seja convertida em servlet e este seja compilado e carregado na memória.

A princípio, não há nada que possa ser feito com JSP que não possa ser realizado com *servlets*. Entretanto, é muito mais conveniente escrever e modificar código HTML comum do que ter muitos comandos para escrita para gerar páginas em HTML. Além disso, separando as tarefas de aparência e conteúdo, pode-se atribuir diferentes tarefas a diferentes pessoas.

2.4.3 JDBC e ODBC

JDBC é uma anacronia para *Java Database Connectivity* que, em português, seria algo parecido com "Conectividade de Banco de Dados em Java". O JDBC constitui um conjunto de classes para integração da linguagem Java com bancos de dados relacionais, sendo uma API (interface para programação de aplicação) que possibilita a adição de comandos SQL (*Structured Query Language*) em uma aplicação Java.

A especificação do JDBC foi baseada no padrão *X/Open SQL Call Level Interface (CLI)*, deixando-o muito parecido com um driver ODBC (*Open Database Connectivity*), que também foi baseado neste padrão. A principal diferença entre estes consiste no fato do ODBC ser uma implementação em linguagem C nativa, enquanto que o JDBC pode ser implementado independentemente de plataforma, seguindo os ideais da linguagem Java.

O objetivo do JDBC é fornecer uma interface independente de qualquer Sistema de Gerenciamento de Banco de Dados (SGBD), permitir um acesso genérico a banco de dados através de SQL e ter uma interface uniforme para diferentes fontes de dados.

Portanto, o programador deve escrever apenas uma interface para interagir com um banco de dados relacional e, utilizando o JDBC, o programa poderá acessar qualquer fonte de dados sem ter uma programação adicional para realizar esta operação.

2.4.4 Web Server

Para utilizar servlets e JSP, é necessário um servidor Web capaz de suportá-los. Na implementação deste trabalho, foi utilizado o Tomcat como servidor.

Tomcat é uma implementação oficial das especificações Servlet 2.2 e JSP 1.1. Pode ser usado como um pequeno servidor "*stand-alone*" para testar servlets e páginas JSP ou pode ser integrado a um servidor Apache. Assim como este, Tomcat é rápido, altamente confiável e grátis. Para maiores detalhes, consultar a *homepage* em <http://jakarta.apache.org/>.

Algumas outras opções para servidores capazes de suportar JSP e servlet são: JavaServer Web Development Kit (JSWDK), Allaire JRun, New Atlanta's ServletExec, Gefion's LiteWebServer (LWS) e Sun's Java Web Server.

CAPÍTULO 3

MODELAGEM E IMPLEMENTAÇÃO DO GERENCIADOR DE ACESSOS

3.1 Considerações iniciais

Basicamente, o objetivo deste trabalho é a criação de um serviço para o registro e geração de estatísticas de acessos, de forma que a coleta de dados seja feita automaticamente no momento em que os visitantes estão acessando um web site. O propósito principal deste serviço é fornecer informações úteis para a análise e gerenciamento.

Para utilizar este serviço, os administradores interessados no registro de acesso de seus usuários deverão cadastrar seu site fornecendo algumas informações a respeito do mesmo. Após o cadastro, os responsáveis terão acesso ao sistema. Assim, poderão alterar informações fornecidas relativas ao site e verificar as estatísticas de acessos e atividades dos usuários registradas pelo sistema.

Uma consideração importante é a preocupação em prover estatísticas detalhadas e precisas, mas que ao mesmo tempo, sejam de fácil leitura e compreensão. Portanto, são usados relatórios estatísticos que facilitem a tomada de decisões e não somente, uma massa de dados confusa e sem valor para aprimorar o empreendimento da empresa. Desta forma, evita-se a frustração dos usuários com o problema da "sobrecarga de informações". Esta ocorre quando o usuário tem muita informação ao seu alcance, mas não tem condições de tratá-la ou de encontrar o que realmente deseja ou lhe interessa [CHE94].

Utilizando este serviço, as empresas poderão sanar questões relativas ao seu web site, como por exemplo:

- Quais são as seções mais acessadas dentro do site?
- Quais são as seções menos acessadas no site?
- Quais os arquivos são mais requisitados para *download*?
- Quais são os dias e horários de maior e menor visitação?
- Quais são os sistemas operacionais e *browsers* utilizados pelos usuários para navegação?

A implementação deste serviço foi realizada de forma a armazenar remotamente os dados capturados. Além disso, todos os programas necessários para geração de estatísticas residem no servidor. Portanto, o uso deste serviço não exige nenhuma programação por parte de seus clientes, nem é necessário instalar qualquer arquivo, programa ou *script* para ter o registro de acessos sendo executado no web site. Só é preciso inserir nas páginas um pequeno código HTML sendo este fornecido quando se cadastra páginas no serviço. Este código, quando colocado num documento HTML, fará uma requisição ao servidor sempre que a página for visitada. Neste processo, informações sobre o visitante do site são extraídas desta requisição e armazenadas em arquivos específicos no servidor.

3.2 Informações sobre o sistema gerado

No sistema gerado, os arquivos de código totalizam 3.695 linhas de código. O sistema é composto por:

- 5 classes em Java;
- 22 arquivos JSP;
- 4 documentos HTML; e
- 2 imagens no formato GIF.

3.3 Diagramas de caso de uso

A Figura 2 apresenta o diagrama de casos de uso do sistema desenvolvido.

O ator denominado Cliente representa o administrador ou responsável pelo Web site. O ator denominado Visitante representa o usuário do site do cliente que terá seu acesso registrado pelo sistema.

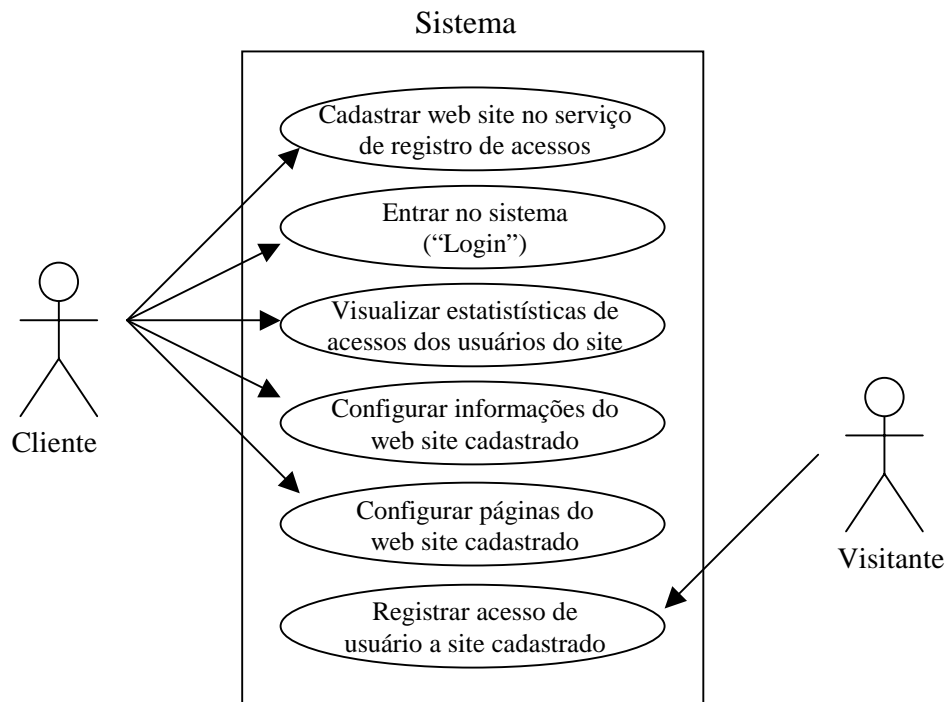


Figura 2 - Diagrama de casos de uso do sistema RAJ

3.4 Diagrama de classes

A figura 3 apresenta o diagrama das classes que foram desenvolvidas na construção do sistema.

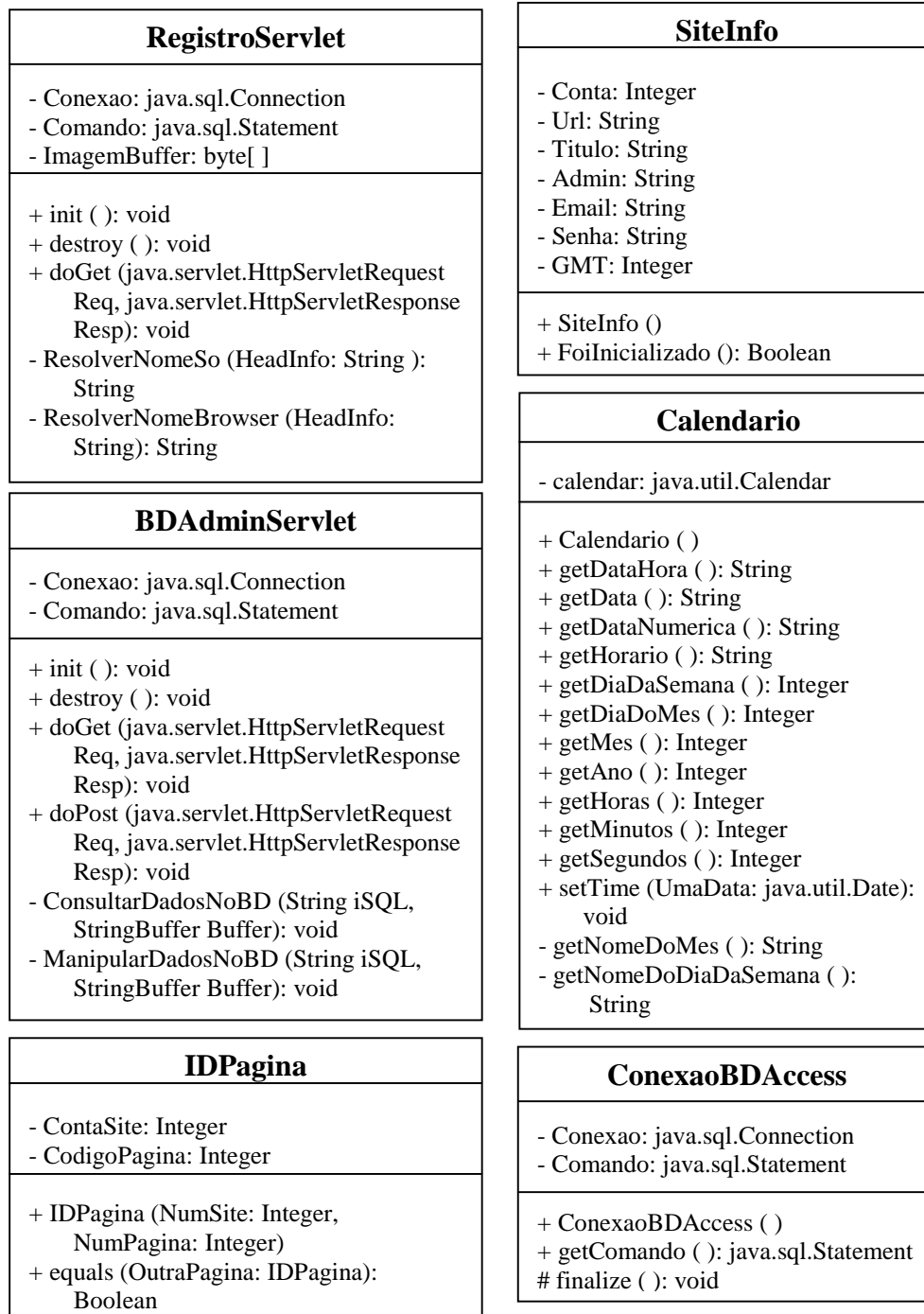


Figura 3 - Diagrama de classes do projeto

CAPÍTULO 4

VALIDAÇÃO DO GERENCIADOR DE ACESSOS

4.1 Instalação e configuração

Desenvolveu-se a aplicação utilizando-se a linguagem Java, a qual permite a instalação e execução do sistema em qualquer plataforma suportada pela máquina virtual da Sun. Apesar de ser multiplataforma e de uso simples, a aplicação exige alguns pré-requisitos para seu correto funcionamento.

Java Development Kit (JDK)

É necessário ter o *kit* de desenvolvimento Java. Além do compilador, gerador de documentação, depurador e outros utilitários para o desenvolvimento de aplicações Java, dispõe de um conjunto de classes que são necessárias para o funcionamento do sistema e para a execução do Tomcat. O JDK pode ser conseguido gratuitamente no site da Sun (<http://www.sun.com>).

Servidor Web capaz de suportar servlets e JSP

É necessário ter instalado um servidor web que suporte essas tecnologias. Durante o desenvolvimento do projeto, foi utilizado o Tomcat, como foi discutido anteriormente. A instalação e configuração deste servidor é bastante simples. Basta fazer o *download* de um arquivo no site oficial (<http://jakarta.apache.org/>) e descompactá-lo em um diretório. As configurações vindas como padrão são suficientes para executar o projeto.

Para iniciar o servidor Tomcat, abra um janela de prompt do DOS. Vá ao diretório onde foi descompactado o Tomcat e execute o *script* de inicialização

do servidor, digitando o comando: "bin\tomcat start". Este *script* ajusta várias configurações automaticamente e inicializa o servidor. Pode ser necessário ajustar a variável de ambiente que indica onde está o JDK. Para isso, utilize o comando "set JAVA_HOME=(dir)", onde (dir) deve ser substituído pelo caminho onde foi instalado o *kit* de desenvolvimento Java.

Para finalizar sua execução, digite o comando "bin\tomcat stop".

Banco de dados

Para armazenar as informações coletadas é necessário ter um Banco de Dados instalado no sistema. O sistema foi desenvolvido e testado em ambiente Windows e utilizou uma fonte de dados ODBC com driver para o Microsoft Access. Por falta de tempo e recursos, não foram testados outros bancos de dados.

RAJ

Atendidos os pré-requisitos, para instalar o sistema só é necessário descompactar o arquivo "raj.zip" dentro do diretório "webapps" que se encontra dentro do diretório onde foi instalado o Tomcat.

4.2 Funcionamento e recursos

O sistema foi desenvolvido para estar disponível como um serviço para a Internet. O responsável ou administrador de web site, interessado em ter o monitoramento dos acessos, deve cadastrar seu site no serviço. Após o cadastro, o cliente fornece informações sobre as páginas cujos acessos deverão ser registrados e, então, o sistema disponibilizará um código em HTML. Este código precisa ser inserido nas páginas do site para que as informações dos usuários sejam capturadas e registradas no momento do acesso.

Depois que o código HTML for inserido nas páginas, os acessos dos usuários serão registrados pelo sistema. A partir destes acessos, estatísticas serão geradas e estarão disponíveis para o cliente. Para visualizar as estatísticas e analisar os dados, o cliente deve ir à *homepage* do serviço e entrar no sistema, fornecendo o número de sua conta e a senha previamente escolhida.

Ao entrar no sistema, o cliente tem acesso a diversos recursos. Pode verificar várias informações a respeito do serviço, configurar informações sobre o web site, cadastrar ou alterar informações sobre páginas monitoradas, além de visualizar as estatísticas de acessos dos seus usuários. Atualmente, o sistema fornece informações estatísticas agrupadas em 5 categorias: número de acessos das páginas registradas, número de visitas organizadas por data, *browsers* utilizados pelos usuários, *hosts* dos usuários do web site e sistemas operacionais utilizados. Algumas telas de demonstração do sistemas podem ser vistas no Apêndice A.

O sistema desenvolvido apresenta algumas características vantajosas, como, por exemplo:

- uso do serviço não exige nenhuma programação por parte de seus clientes;
- não é necessário instalar qualquer arquivo para ter o registro de acessos sendo executado no web site;
- qualquer pessoa que tenha seu site na Internet, hospedado em servidor público ou dedicado, pode fazer uso do serviço;
- informações sempre atualizadas sobre os acessos dos usuários ficam disponíveis *online* para serem consultadas em qualquer lugar onde haja um computador conectado à Internet.

Algumas deficiências apresentadas pelo serviço necessitariam de um tempo extra para serem corrigidas. Como exemplo destas, pode-se mencionar o pequeno número de informações fornecidas a respeito do usuário. Um maior

estudo seria necessário para que houvesse um melhor aproveitamento dos dados obtidos, gerando assim não somente informações estatísticas que refletem uma contagem dos dados armazenados como também, inferências a respeito do comportamento navegacional dos usuários para um gerenciamento mais eficiente do site. Também, não houve implementação da apresentação dos dados estatísticos utilizando recursos gráficos, tais como, gráficos de barras, linhas ou áreas. Estes recursos certamente facilitariam a visualização dos resultados da interação dos usuários. Além disso, a captura de algumas informações desejadas como, por exemplo, se o navegador do usuário permite ou não o recebimento de *cookies*, se a execução de Java e JavaScript estão habilitadas, e qual o documento anteriormente acessado, não foi possível devido a limitações das informações disponíveis nos cabeçalhos das requisições HTTP e também, devido a problemas com a utilização de JavaScript no código HTML fornecido ao cliente.

CAPÍTULO 5

CONCLUSÕES E TRABALHOS FUTUROS

Neste estudo, discutiu-se a necessidade de informações a respeito dos usuários para análise e organização de um web site. Usando diversos exemplos, uma discussão foi apresentada sobre a utilidade das informações para estruturar e projetar um site adaptativo. Apresentou-se também um modelo para o processo de personalização de web sites e, na investigação sobre formas de coleta de dados sobre usuários, mostrou-se dois métodos possíveis para captura de informações úteis para a administração de um empreendimento virtual.

Uma ferramenta foi desenvolvida em forma de serviço para a Internet, com o intuito de capturar dados e gerar estatísticas sobre usuários de web sites. Sua implementação foi comentada como também as tecnologias envolvidas na criação do serviço. Embora o sistema disponha de poucos recursos, fornece uma base inicial para uma aplicação que pode se tornar muito útil no gerenciamento de web sites.

Diversas melhorias precisam ser implementadas para que o sistema forneça aos administradores mais recursos para uma melhor compreensão do comportamento dos usuários e assim, gerenciar melhor seu empreendimento. Por exemplo, não utilizar somente relatórios, mas também recursos estatísticos gráficos para uma melhor visualização e entendimento das estatísticas obtidas sobre os usuários. Além disso, para um melhor aproveitamento dos dados capturados, utilizar algoritmos de recuperação de informações e técnicas de mineração de dados a fim de obter inferências a respeito do comportamento navegacional dos usuários.

CAPÍTULO 6

REFERÊNCIAS BIBLIOGRÁFICAS

- [CHA95] CHAKRAVARTHY, A. S.; HAASE, K. B. **NetSerf: using semantic knowledge to find Internet information archives.** Proceedings. SIGIR, 1995.
- [SSU96] SILBERSCHATZ, A.; STONEBRAKER, M.; ULLMAN, J. **Database research: achievements and opportunities into the 21st century.** Journal of the Brazilian Computer Society, v.2, n.3, 1996.
- [TYG97] TYGAR, J. D. **Atomicity in Electronic Commerce, In Internet Besieged.** Addison-Wesley and ACM Press, 1997
- [AIS93] AGRAWAL , R.; IRNIELINKSI, T.; SWARNI, A. **Mining Association Rules Between Sets of itens in large databases.** In. Proc. 1993 ACM-sigmod Int. Conf. Management of Data, 1993
- [GRE00] GREENING, D. R. **Data mining on the web.** Web Techniques, 2000.
- [PET97] PERKOWITZ, M.; ETZIONI, O. **Adaptative web sites: Automatically learning from user access patterns.** Technical Report UW-CSE-9703-01, University of Washington, Department of Computer Science and Engineering, 1997.
- [RRF99] RUAS, F.; RIBEIRO, F.; FREITAS, L. **Técnica para a construção de web sites adaptativos.** Technical report, Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, 1999
- [FKN96] FINK, J.; KOBASA, A.; NILL, A. **User-oriented adaptivity and adaptability in avanti project.** In Designing for the Web: Empirical Studies. Microsoft Usability Group, 1996.

- [FSS99] FU, Y.; SANDHU, K.; SHIH, M. **Clustering of web users based on access patterns.** Workshop on Web Usage Analysis and User Profiles, 1999.
- [GHL00] GOMORY, S.; HOCH, R.; LEEA. et al. **E-commerce intelligente measuring, analyzing and reporting on merchandising effectiveness of online stores.** IBM Institute for Advanced Commerce at <http://www.ibm.com/iac/tech-paper.html>, 2000
- [JFM97] JOACHIMS, T.; FREITAG, D.; MITCHELL, T. **Webwatcher: a tour guide for the world wide web.** 15th Int Joint Conf, 1997.
- [MDU99] MURRAY, D.; DURRELL, K. **Inferring demographic attributes of anonymous internet users.** Workshop on Web Usage Analysis and User Profiles, 1999.
- [PET99] PERKOWITZ, M.; ETZIONI, O. **Towards adaptative web sites: Conceptual framework and case study.** pages 167-180. Proc. 8th Int. World Wide Web Conf., 1999.
- [PET98] PERKOWITZ, M.; ETZIONI, O. **Adaptative web sites: Automatically synthesizing web pages.** Proc. 15th Nat. Conf. On Artificial Intelligence, 1998.
- [BBA99] BUCHNER, A. G.; BAIRMGARTEN, M.; Anand, S. S. et al **Navigation pattern discovery from internet data.** Workshop on Web Usage Analysis and User Profiles, 1999.
- [CHA99] CHAN, P. K. **A non-invasive learning approach to building web users profiles.** Workshop on Web Usage Analysis and User Profiles, 1999.
- [TMU00] TAO, F.; MURTAGH, F. **Towards knowledge discovery from WWW log data.** In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'00). IEEE, Las Vegas, Nevada, 2000
- [ILU99] ILUX EDGE E-BUSINESS COMPANY. **ilux Edge** URL: <http://www.ilux.com/>, 1999

- [MAR98] MARKETWAVE CORPORATION. **Marketwave** URL: <http://www.marketwave.com/>, 1998
- [BAR01] BARRETT, Bradford L. **Webalizer** URL: <http://www.webalizer.com/>, 2001
- [WEB99] WEBTRENDS CORPORATION. **Webtrends** URL: <http://www.webtrends.com/>, 1999
- [HIG01] HIGHWIRED INTERNET INNOVATIONS INC. **HWCounter** URL: <http://www.hwcounter.com/>, 2001
- [MYC01] MYCOMPUTER, INC. **SuperStats** URL: <http://www.counter.com/>, 2001
- [MAS01] MASTERS., AARON **IDStat** URL: <http://www.idstat.com/>, 2001
- [ROO01] ROOTS, KURT **Jinko** URL: <http://www.jinko.com/counters/>, 2001
- [WAT01] WATCHWISE, INC. **WatchWise** URL: <http://www.watchwise.com/>, 2001
- [CHE94] CHEN, H. **A textual database/knowledge-base coupling approach to creating computer-supported organizational memory.** MIS Department, University of Arizona, 1994.
- [TAK00] TAKAHASHI, T. (org) **Sociedade da Informação no Brasil: livro Verde** Brasília: Ministério da Ciência e Tecnologia, 2000
- [FEI91] FEIGENBAUM, A. V. **Total quality control** 3.ed., McGraw-Hill, 1991
- [HAL00] HALL, M. **Core Servlets and JavaServer Pages** Sun Microsystems Press and Prentice Hall, 2000

APÊNDICE A

AMOSTRAS DE TELAS DO SISTEMA

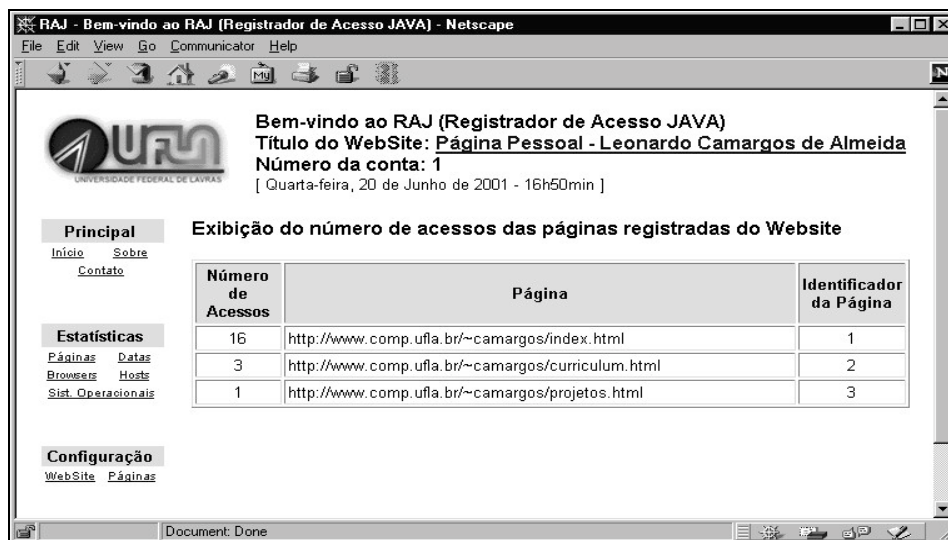


Figura 4 - Amostra da tela de exibição do número de acesso das páginas registradas de um web site

Sistemas operacionais utilizados pelos usuários do Website		
Número de Visitas	Percentagem	Sistema Operacional
2	11.764706%	Linux
15	88.23529%	Windows NT
Total de Visitas:17		

Figura 5 - Amostra de parte da tela de exibição de sistemas operacionais utilizados pelos usuários de um web site

Browsers utilizados pelos usuários do Website

Número de Visitas	Porcentagem	Navegador
3	17.647058%	Internet Explorer 5.5
1	5.882353%	Netscape Navigator 4.6
11	64.70588%	Netscape Navigator 4.75
2	11.764706%	Netscape Navigator 4.76
Total de Visitas:17		

Figura 6 - Amostra de parte da tela de exibição de browsers utilizados pelos usuários de um web site

Hosts dos usuários do Website

Número de Visitas	Porcentagem	Host dos usuários
1	5.882353%	192.168.100.22
2	11.764706%	192.168.100.40
12	70.588234%	gwlab.comp.ufla.br
2	11.764706%	localhost
Total de Visitas:17		

Figura 7 - Amostra de parte da tela de exibição dos hosts dos usuários

Configuração de Informações sobre o Website

Número da conta: 1

Endereço (URL) do Website:

Nome ou Título do Website:

Endereço de correio eletrônico (Email):

Nome do Administrador:

Senha:

Data e Hora Local: ▼

Figura 8 - Amostra de parte da tela de configuração de informações do web site

Páginas Registradas do Website: <http://www.comp.ufla.br/~camargos/>

Identificador	URL relativa	Título da Página	Código HTML
1	index.html	Página Principal	Código HTML
2	curriculum.html	Curriculum Vitae	Código HTML
3	projetos.html	Projetos desenvolvidos	Código HTML

Figura 9 - Amostra de parte da tela de exibição de páginas monitoradas pelo serviço de registro de acesso

Incluir Registro de Página

URL da Página:

Título da Página:

Figura 10 - Amostra de parte da tela de cadastro de página a ser monitorada pelo serviço de registro de acessos