

**MARCELLO VERONESE**



**FERRAMENTA DE SUPORTE À  
NAVEGAÇÃO EM PORTAIS CORPORATIVOS  
BASEADA NO COMPORTAMENTO DO  
USUÁRIO**

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para a obtenção do título de Mestre em Informática Aplicada.

**CURITIBA**

**2002**

Veronese, Marcello Kieling

Ferramenta de suporte à navegação em Portais Corporativos baseada no comportamento do usuário. Curitiba, 2002. 6p.

Dissertação – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada.

1.Personalização de Site 2.Portal Corporativo 3.Ferramenta de Suporte à navegação 4.Comportamento do usuário *web*. I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Informática Aplicada II.Ferramenta de suporte à navegação em Portais Corporativos baseada no comportamento do usuário



Pontifícia Universidade Católica do Paraná  
Centro de Ciências Exatas e de Tecnologia  
Programa de Pós-Graduação em Informática Aplicada

ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO  
DO PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA  
DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

DEFESA DE DISSERTAÇÃO Nº 52

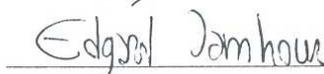
Aos 21 dias do mês de março de 2002 realizou-se a sessão pública de defesa da dissertação "**Ferramenta de Suporte à Navegação em Portais Corporativos Baseada no Comportamento do Usuário**", apresentada por **Marcello Kieling Veronese** como requisito parcial para a obtenção do título de **Mestre em Ciências**, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Carlos Maziero  
PUCPR (Presidente)

  
assinatura

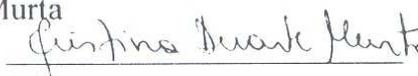
APROVADO  
parecer

Prof. Dr. Edgard Jamhour  
PUCPR



APROVADO

Profa. Dra. Cristina Duarte Murta  
UFPR

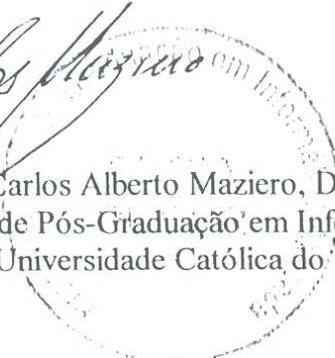
  
APROVADO

Conforme as normas regimentais do PPGIA e da PUCPR, o trabalho apresentado foi considerado APROVADO SEM RESTRIÇÕES (aprovado sem restrições, aprovado com exigências ou reprovado), segundo a avaliação da maioria dos membros da Banca Examinadora acima indicada.

Observações, exigências e/ou restrições da Banca Examinadora, quando houver:

NADA A RELATAR

continuar no verso, se necessário

Prof. Dr. Carlos Alberto Maziero, Diretor  
Programa de Pós-Graduação em Informática Aplicada  
Pontifícia Universidade Católica do Paraná

A meus Pais, origem de tudo.

A meus filhos, futuro.

À Giuliana, incentivo no estudo.

Meus Amores.

## Agradecimentos

Diversas formas de agradecimento podem ser idealizadas e sonhadas, pensadas e acalentadas, concebidas e desejadas, todas, porém, com um defeito em comum: não traduzem a efetividade das emoções contidas. Difícil buscar palavras para agradecer aos Mestres que incentivaram e desafiaram, buscando nossa superação. Difícil encontrar sentido para o desafio de vida que exigiram nossas forças. Difícil expressar a dor de momentos de solidão que cobraram nossas vidas. Por todos estes, e outros tantos, é que se deve agradecer. Acima de tudo olhando para frente e buscando prosseguir, levantando os olhos e buscando forças novas. Algumas breves palavras de incentivo têm este poder e realmente surtem efeito. Aos colegas de Mestrado que, mesmo num efêmero momento, formaram minha família. Aos professores que instigaram minha imaginação. E, sobretudo, ao meu orientador Prof. Dr. Carlos Maziero, que apostou em meu potencial apesar de todos os problemas e dificuldades por mim enfrentados: Obrigado!

## Sumário

<b>Agradecimentos</b>	ix
<b>Sumário</b>	xi
<b>Lista de Figuras</b>	xiv
<b>Lista de Tabelas</b>	xv
<b>Lista de Símbolos</b>	xvi
<b>Lista de Abreviaturas</b>	xvii
<b>Resumo</b>	xix
<b>Abstract</b>	xxi

### Capítulo 1

<b>INTRODUÇÃO</b>	<b>01</b>
1.1. Paradigma da Conectividade.....	03
1.2. Contexto da Análise.....	03
1.3. Desafios da informação no âmbito da web.....	05
1.4. Foco de Avaliação.....	07
1.5. Organização.....	09

### Capítulo 2

<b>WEB NO CONTEXTO DAS CORPORAÇÕES</b>	<b>13</b>
2.1. A Web no Ambiente Corporativo.....	13
2.2. Soluções standard x Soluções caseiras.....	14
2.3. Características do Portal Corporativo.....	15
2.4. Portal Business-to-Employee (B2E).....	17

2.5. Gestão do Conhecimento.....	18
2.6. Conclusão.....	22
<b>Capítulo 3</b>	
<b>PORTAL CORPORATIVO - ESQUEMA CONCEITUAL</b>	<b>23</b>
3.1. Definição de Portal Corporativo.....	25
3.2. Portais Públicos ou Corporativos.....	29
3.2.1. Portal Público.....	30
3.2.2. Portal Corporativo.....	31
3.3. Tipos de Portais Corporativos.....	32
3.3.1. Portal de Informações.....	32
3.3.2. Portal de Apoio à Decisão.....	33
3.3.2.1. Portal de Apoio à Decisão baseado em ferramenta de business-intelligence.....	33
3.3.3. Portal de Negócios.....	34
3.3.4. Portal Cooperativo.....	34
3.3.4.1. Portal Cooperativo baseado em ferramenta de cooperação.....	34
3.3.5. Portal de Integração de Especialistas.....	35
3.3.6. Portal de Informações Empresariais.....	35
3.3.7. Portal do Conhecimento.....	36
3.3.8. Visão Geral dos Tipos de Portais Corporativos.....	36
3.4. Requisitos mínimos de um Portal Corporativo.....	37
3.5. Conclusão.....	40
<b>Capítulo 4</b>	
<b>ANÁLISE DE SOLUÇÕES</b>	<b>41</b>
4.1. Customização de sites.....	41
4.2. Data Webhousing.....	43
4.3. Web Mining.....	45
4.3.1. Mineração de Conteúdo da Web.....	46
4.3.2. Mineração de Uso da Web.....	47
4.4. Ferramentas de Mercado.....	47
4.4.1. MyLivelink Portal.....	48

4.4.2. Plumtree Corporate Portal 4.0.....	48
4.4.3. K-Station.....	49
4.4.4. Broadvision.....	50
4.4.5. Websphere Personalization.....	50
4.5. Conclusão.....	51
<b>Capítulo 5</b>	
<b>SUORTE A NAVEGAÇÃO BASEADO NO PERFIL DO USUÁRIO</b>	<b>53</b>
5.1. Arquitetura.....	54
5.2. Requisitos a serem contemplados.....	55
5.3. Resultados Esperados pelo Modelo Proposto.....	57
5.4. Visão Geral da Solução.....	58
5.5. Projeto Físico.....	59
5.6. Modelo Conceitual.....	61
5.6.1. Modelo de Dados.....	62
5.6.2. Esquema Geral.....	63
5.6.3. Processamento dos Grafos.....	64
5.6.4. Implementação Computacional.....	65
5.6.5. Especificação dos Processo.....	66
5.6.6. Determinação do Período de Inatividade Máximo.....	68
5.6.7. Limitações.....	69
5.7. Resultados Obtidos.....	70
5.8. Conclusão.....	73
<b>CONCLUSÃO</b>	<b>75</b>
<b>Referências Bibliográficas</b>	<b>79</b>
<b>Glossário</b>	<b>83</b>
<b>Apêndice A - Unidade Básica de Referência de Navegação</b>	<b>85</b>

## Lista de Figuras

Figura 3.1	Tipos de Portais baseado na utilização.....	36
Figura 4.1	Distribuição de Fornecedores avaliando Customização [ZUL01].....	42
Figura 5.1	Visão geral do processo de suporte a navegação em um Portal Corporativo.....	59
Figura 5.2	Modelo Conceitual.....	61
Figura 5.3	Modelo Entidade-Relacionamento.....	62
Figura 5.4	Esquema Geral.....	63

## Lista de Tabelas

Tabela 1.1	Conceitos Internet, Intranet e Extranet	8
Tabela 2.1	Características de um Portal Corporativo	16
Tabela 3.1	Gerações dos Portais Públicos	30
Tabela 3.2	Gerações dos Portais Corporativos	32
Tabela 3.3	Requisitos mínimos de um portal corporativo	39
Tabela 5.1	Requisitos mínimos da ferramenta de suporte à navegação	56
Tabela 5.2	Requisitos não contemplados	57
Tabela 5.3	Principais tabelas	62
Tabela 5.4	Resultados de navegação - Modo Convencional	71
Tabela 5.5	Consolidado navegação - Modo Convencional	72
Tabela 5.6	Resultados de navegação - Modo apoiado Ferramenta	72
Tabela 5.7	Consolidado navegação - Modo apoiado Ferramenta	72
Tabela 5.8	Resultados Alcançados	73

## Lista de Símbolos

$\Sigma$	Somatório de $n$ elementos
$(a_1, a_n)$	Elementos inicial e final de uma amostra
$N$	Número de elementos de uma amostra
$S = (P_S, L_S)$	Grafo definido como um par de vértices e arestas
$P_S$	Conjunto finito de vértices
$L_S$	Conjunto finito de arestas onde $A_G \subseteq (V_G / 2)$
$\Delta_{\max}$	Intervalo de tempo máximo entre duas páginas em uma navegação no site

## Lista de Abreviaturas

HTTP	<i>Hipertext Transfer Protocol</i>
WWW	<i>World Wide Web</i>
W3C	<i>Web Consortium</i>
TCP/IP	<i>Transfer Control Protocol / Internet Protocol</i>
HTML	<i>Hiper Text Markup Language</i>
OCR	<i>Optical Character Recognition</i>
ERP	<i>Enterprise Resource Planning</i>
OLAP	<i>On-line Analytical Processing</i>
EIP	<i>Enterprise Information Portal</i>
DSS	<i>Decision Support Systems</i>
XMI	<i>Extensible Markup Language</i>
API	<i>Application Programming Interface</i>
J2EE	<i>Java TM 2 Platform Enterprise Edition</i>
CGI	<i>Common Gateway Interface</i>
CRM	<i>Customer Relationship Management</i>
SGML	<i>Standard Generalized Markup Language</i>

## Resumo

A Internet tem se estabelecido como o mais importante fenômeno do final do século XX, com a possibilidade de virtualmente conectar pessoas ou corporações em qualquer lugar do mundo, baseado em uma interface simples, padronizada e de baixo custo para todos os envolvidos. Nas corporações a Internet é apresentada com um nome específico: Intranet. É notória a influência da Intranet nas corporações pois é possível integrar informações e aplicações sob um único canal de apresentação que representa um autêntico “*Desktop Virtual*” para o funcionário.

Neste contexto, a gestão do conhecimento torna-se fator decisivo para o aumento da produtividade do funcionário, evitando desperdício de tempo na busca de informações corporativas. Assim, a questão principal deste trabalho é visada: Como apresentar um único ponto de acesso a todas as informações da organização sem gerar trabalho adicional tornando o empregado um profissional não eficiente? A resposta a este ponto é o uso de um Portal Corporativo apoiado por uma ferramenta de suporte à navegação baseada no comportamento do usuário web.

Esta dissertação apresenta o paradigma da conectividade onde a web é o centro da referência do trabalho, contextualizando a evolução ocorrida na indústria do software. Os principais desafios da informação são identificados no âmbito da web, tais como: distribuição dos dados, alta percentagem de volatilidade, volume elevado de informações, dados redundantes e não-estruturados, qualidade e a heterogeneidade dos dados. É apresentado o foco de avaliação que é dedicado ao Portal Corporativo, notadamente de instituições financeiras, e a proposta de apoiar a navegação dos usuários a partir da análise dos registros de log dos servidores web.

## Resumo

A Internet tem se estabelecido como o mais importante fenômeno do final do século XX, com a possibilidade de virtualmente conectar pessoas ou corporações em qualquer lugar do mundo, baseado em uma interface simples, padronizada e de baixo custo para todos os envolvidos. Nas corporações a Internet é apresentada com um nome específico: Intranet. É notória a influência da Intranet nas corporações pois é possível integrar informações e aplicações sob um único canal de apresentação que representa um autêntico “*Desktop Virtual*” para o funcionário.

Neste contexto, a gestão do conhecimento torna-se fator decisivo para o aumento da produtividade do funcionário, evitando desperdício de tempo na busca de informações corporativas. Assim, a questão principal deste trabalho é visada: Como apresentar um único ponto de acesso a todas as informações da organização sem gerar trabalho adicional tornando o empregado um profissional não eficiente? A resposta a este ponto é o uso de um Portal Corporativo apoiado por uma ferramenta de suporte à navegação baseada no comportamento do usuário web.

Esta dissertação apresenta o paradigma da conectividade onde a web é o centro da referência do trabalho, contextualizando a evolução ocorrida na indústria do software. Os principais desafios da informação são identificados no âmbito da web, tais como: distribuição dos dados, alta percentagem de volatilidade, volume elevado de informações, dados redundantes e não-estruturados, qualidade e a heterogeneidade dos dados. É apresentado o foco de avaliação que é dedicado ao Portal Corporativo, notadamente de instituições financeiras, e a proposta de apoiar a navegação dos usuários a partir da análise dos registros de log dos servidores web.

Fica demonstrado que as funções mais importantes de um Portal Corporativo são: Gestão do Conhecimento, Apoio à Decisão e Processamento Cooperativo. Conforme apresentado neste documento, o uso de certas nomenclaturas pode levar a interpretações equivocadas sobre as reais características de um Portal. Para esclarecer esta questão um padrão dos conceitos é providenciado a partir da avaliação dos principais relatórios.

A capacidade de facilitar o acesso dos usuários às informações institucionais está intrinsecamente relacionada à uma abordagem amigável. Para atingir este benefício é fundamental que o projeto do Portal Corporativo considere a interação dos usuários com sua interface e estabeleça como objetivo minimizar seus problemas de usabilidade. A proposta de uma ferramenta de suporte ao acesso às informações disponíveis em um Portal Corporativo baseada na avaliação do perfil de navegação do usuário (característica de usabilidade) será útil para qualquer instituição que mantenha informações sob a forma digital. Este trabalho mostra que é possível rastrear o comportamento de usuários dentro de um site de uma Intranet, notadamente num Portal Corporativo. Este comportamento, ou perfil, pode ser modelado como um grafo e pode ser armazenado num banco de dados de conhecimento para ser usado como referência por outro usuário.

Neste trabalho, foi utilizado um *servlet* que permite personalizar o site. Um conjunto de páginas sugeridas é apresentado para o usuário durante a navegação tendo como base o perfil do grupo onde ele é membro. Estas sugestões irão caracterizar um ganho de produtividade pela diminuição no tempo de busca de informações. Entretanto, para montar o modelo de comportamento foi necessário assumir algumas premissas e heurísticas para determinar o relacionamento entre duas páginas. Ainda assim, o problema em sua essência é resolvido pelo uso de uma ferramenta de suporte à navegação baseada no comportamento do usuário web num Portal Corporativo. No futuro, outros trabalhos poderão usar alguns mecanismos mais complexos que poderão incrementar esta ferramenta tornando possível dispensar estas premissas e determinar mais eficientemente os relacionamentos entre as páginas.

**Palavras-chave:** 1. Personalização de Site 2. Portal Corporativo 3. Ferramenta de Suporte à navegação 4. Comportamento do usuário *web*.

Fica demonstrado que as funções mais importantes de um Portal Corporativo são: Gestão do Conhecimento, Apoio à Decisão e Processamento Cooperativo. Conforme apresentado neste documento, o uso de certas nomenclaturas pode levar a interpretações equivocadas sobre as reais características de um Portal. Para esclarecer esta questão um padrão dos conceitos é providenciado a partir da avaliação dos principais relatórios.

A capacidade de facilitar o acesso dos usuários às informações institucionais está intrinsecamente relacionada à uma abordagem amigável. Para atingir este benefício é fundamental que o projeto do Portal Corporativo considere a interação dos usuários com sua interface e estabeleça como objetivo minimizar seus problemas de usabilidade. A proposta de uma ferramenta de suporte ao acesso às informações disponíveis em um Portal Corporativo baseada na avaliação do perfil de navegação do usuário (característica de usabilidade) será útil para qualquer instituição que mantenha informações sob a forma digital. Este trabalho mostra que é possível rastrear o comportamento de usuários dentro de um site de uma Intranet, notadamente num Portal Corporativo. Este comportamento, ou perfil, pode ser modelado como um grafo e pode ser armazenado num banco de dados de conhecimento para ser usado como referência por outro usuário.

Neste trabalho, foi utilizado um *servlet* que permite personalizar o site. Um conjunto de páginas sugeridas é apresentado para o usuário durante a navegação tendo como base o perfil do grupo onde ele é membro. Estas sugestões irão caracterizar um ganho de produtividade pela diminuição no tempo de busca de informações. Entretanto, para montar o modelo de comportamento foi necessário assumir algumas premissas e heurísticas para determinar o relacionamento entre duas páginas. Ainda assim, o problema em sua essência é resolvido pelo uso de uma ferramenta de suporte à navegação baseada no comportamento do usuário web num Portal Corporativo. No futuro, outros trabalhos poderão usar alguns mecanismos mais complexos que poderão incrementar esta ferramenta tornando possível dispensar estas premissas e determinar mais eficientemente os relacionamentos entre as páginas.

**Palavras-chave:** 1. Personalização de Site 2. Portal Corporativo 3. Ferramenta de Suporte à navegação 4. Comportamento do usuário *web*.

## Abstract

The Internet has been established as the most important phenomenon of the end of XX century allowing virtual connection of people and corporations in any place of the world, based on simple, standard and low cost interface to all involved. At Corporations the Internet is presented with a specific name: Intranet. It is well-known the influence of Intranet at Corporations because it makes possible to integrate information and applications above only one presentation channel that represents an authentic “*Virtual Desktop*” to employee.

At this context, the knowledge management became a decisive factor to improve the productivity of employee avoiding time-wasting by search corporate information. In this way, the main question of this work is achieved: How to make possible to present only one access point to all corporate information without to generate additional work becoming employee in non-efficient professional? The answer to this question is the use of a Corporate Portal structured above a navigation support tool based on web user behaviour.

This dissertation presents the connectivity paradigm where web is the core reference of the study by evaluating the context of software industry evolution. The main challenges of information is identified at web scope such as: data distribution, high percentage of volatility, huge volume of information, redundant and non-structured data, quality of data and heterogeneous sources. It is presented the focus of evaluation that is dedicated to Corporate Portal - mainly financial corporations - and the proposal to support the navigation of users by analysis from log records of web servers.

It was demonstrated that the most important functions of a Corporate Portal are: Knowledge Management, Decision Support and Cooperative Processing. According to this document the use of certain nomenclatures can direct to wrong interpretations about the

## Abstract

The Internet has been established as the most important phenomenon of the end of XX century allowing virtual connection of people and corporations in any place of the world, based on simple, standard and low cost interface to all involved. At Corporations the Internet is presented with a specific name: Intranet. It is well-know the influence of Intranet at Corporations because it make possible to integrate information and applications above only one presentation channel that represents an authentic "*Virtual Desktop*" to employee.

At this context, the knowledge management became a decisive factor to improve the productivity of employee avoiding time-wasting by search corporate information. In this way, the main question of this work is achieved: How to make possible to present only one access point to all corporate information without to generate additional work becaming employee in non-efficient professional? The answer to this question is the use of a Corporate Portal structured above a navigation support tool based on web user behaviour.

This dissertation presents the connectivity paradigm where web is the core reference of the study by evaluating the context of software industry evolution. The main challenges of information is identified at web scope such as: data distribution, high percentage of volatility, huge volume of information, redundant and non-structured data, quality of data and heterogeneous sources. It is presented the focus of evaluation that is dedicated to Corporate Portal - mainly financial corporations - and the proposal to support the navigation of users by analysis from log records of web servers.

It was demonstrated that the most important functions of a Corporate Portal are: Knowledge Management, Decision Support and Cooperative Processing. According to this document the use of certain nomenclatures can direct to wrong interpretations about the

correct Portal characteristics. To clear this issue a standard of concepts was provided by evaluation of main reports.

The capability to make easy the users access to corporate information is directly related to user-friendly approach. To aim this benefit is fundamental that the design of Corporate Portal considers the users interactions with their interfaces and establishes the objective of minimizing their usability problems. The proposal of an access support tool to information that is available at Corporate Portal based on the evaluation of user navigation profile (usability characteristic) will be useful to any corporation that maintain information in digital format. This work demonstrates that it is possible to track the user behaviour at Intranet site, mainly at Corporate Portal. This web user behaviour or profile can be drawn as a graph and can be recorded at knowledge database to be used like reference by other user.

In this work was used a *servlet* that make possible to customize the site. A set of suggested pages is presented to the user during the navigation process based on the group profile where it is a member. These suggestions will represent a productivity improvement by reduction time of retrieval information. However, to make behaviour Model it was necessary to assume some premises and heuristics to determine the relationship between two pages. Despite this, the problem at its essential was solved by the use of a navigation support tool based on web user behaviour at Corporate Portal. In the future other jobs could use some more complex mechanisms that would improve this tool making possible to dismiss this premises and to determine more efficiently the relationship between pages.

**Keywords:** 1. Site personalization 2. Corporate Portal 3. Navigation support tool 4. Web user behaviour.

correct Portal characteristics. To clear this issue a standard of concepts was provided by evaluation of main reports.

The capability to make easy the users access to corporate information is directly related to user-friendly approach. To aim this benefit is fundamental that the design of Corporate Portal considers the users interactions with their interfaces and establishes the objective of minimizing their usability problems. The proposal of an access support tool to information that is available at Corporate Portal based on the evaluation of user navigation profile (usability characteristic) will be useful to any corporation that maintain information in digital format. This work demonstrates that it is possible to track the user behaviour at Intranet site, mainly at Corporate Portal. This web user behaviour or profile can be drawn as a graph and can be recorded at knowledge database to be used like reference by other user.

In this work was used a *servlet* that make possible to customize the site. A set of suggested pages is presented to the user during the navigation process based on the group profile where it is a member. These suggestions will represent a productivity improvement by reduction time of retrieval information. However, to make behaviour Model it was necessary to assume some premises and heuristics to determine the relationship between two pages. Despite this, the problem at its essential was solved by the use of a navigation support tool based on web user behaviour at Corporate Portal. In the future other jobs could use some more complex mechanisms that would improve this tool making possible to dismiss this premises and to determine more efficiently the relationship between pages.

**Keywords:** 1. Site personalization 2. Corporate Portal 3. Navigation support tool 4. Web user behaviour.

# Capítulo 1

## INTRODUÇÃO

A Internet vem se constituindo no grande fenômeno deste final de século, pela possibilidade de conectar, virtualmente, pessoas ou organizações em qualquer ponto do mundo, de forma simples, padronizada, e com baixo custo a todos os envolvidos.

Desde os primeiros momentos em que começou a ser divulgada, alguns visionários fizeram ousadas previsões sobre a onda de mudanças que a Internet iria causar, pois entendiam que diversas barreiras estavam literalmente quebradas, como distância geográfica, acesso atualizado a informações, hábitos e mercados, tudo num clicar de botões. Algumas projeções ficaram até bastante aquém do previsto, pois o número de equipamentos ligados à rede, quer sejam servidores ou usuários, ultrapassou em muito os números mais otimistas.

Em alguns países, como o Brasil, o ritmo de crescimento é muitas vezes maior que a média mundial. O uso comercial da Internet evoluiu bastante desde as primeiras e tímidas páginas de divulgação institucional, e mostra sinais de ter encontrado o caminho e o ritmo certos:

- O acesso a bases de dados reais das empresas, para consulta a estoques, ordens de produção, registro de propostas de seguros, acompanhamento da entrega de produtos tem sido cada vez mais utilizado através da Internet;
- Transações bancárias efetivadas pela Internet, têm apresentado nível de utilização bastante acima das previsões mais otimistas;

- O uso crescente e constante da Internet para as compras em varejos específicos como livrarias, supermercados, lojas de discos;
- Crescimento expressivo do uso de intranets, redes corporativas de computadores que utilizam a infra-estrutura da Internet para sua interconexão, incluindo o acesso à base de dados da organização para o público externo às corporações.

Não há dúvidas que o conteúdo da Internet atualmente excede o de qualquer biblioteca do mundo em informação, e se tornará rapidamente nos próximos dez anos a principal fonte de conhecimento usada pela mente humana. O "boom" da utilização da Internet, em seu contexto globalizado conhecido como World Wide Web, tem seu início efetivo a partir do final dos anos 80 com um crescimento exponencial agora bem conhecido. A título de referência, apenas o conjunto de informações textuais atualmente disponíveis são estimados além da ordem do terabyte. Além disso, outras mídias, tais como imagens, áudio e vídeo também estão disponíveis.

Assim sendo, a Internet pode ser vista como um grande banco de dados não estruturado e distribuído. Esta característica aponta para a necessidade de disponibilização de ferramentas de gerenciamento, recuperação e filtro de informações deste mesmo banco de dados. Este problema é especialmente importante em extensas Intranets de onde se desejam extrair ou inferir novas informações para suporte ao processo de decisões. Por analogia dos modelos de data-warehouse, a extração e inferência de novas informações a partir da Internet podem ser denominadas como "*web mining*".

O conhecimento é atualmente um dos maiores ativos de uma instituição. Na indústria financeira esta afirmação adquire proporções significativas e requer a definição de um padrão de pesquisa. Sob o ponto de vista das instituições, esta necessidade tem sido parcialmente contemplada pelo emprego de Portais Corporativos, cuja proposta é facilitar o acesso às informações digitais no contexto organizacional.

Este trabalho conceitua o emprego de Portais Corporativos, contemplando as diversas classificações conforme as facilidades apresentadas, focando no objetivo geral de possibilitar a análise do perfil de navegação do usuário e a correspondente adequação do Portal a cada perfil determinado.

O foco deste trabalho é analisar os dados armazenados pela log dos servidores web visando possibilitar análises comportamentais dos usuários, com a intenção de determinar o conjunto de informações que mais interessam para a organização e privilegiando a divulgação desse acesso, bem como atingir o decorrente aumento de produtividade por conta da maior eficiência no processo de busca de informações.

### **1.1. Paradigma da Conectividade**

Nos últimos anos, o impacto da Web originou um novo paradigma de conectividade. A Web tornou-se um meio de acesso e comunicação mundialmente abrangente, relativamente seguro, a cada dia torna-se mais interativo e, sobretudo barato.

A disseminação desta nova tecnologia não ficou restrita somente a grandes corporações, centros de pesquisa e universidades, ela tem se espalhado por todos os níveis da sociedade, interconectando todos sem distinções. A velocidade com que a Web foi adotada não possui nenhum paralelo quando comparada com a adoção de qualquer meio de comunicação, como por exemplo, jornal, rádio ou televisão.

O uso da tecnologia Web por grande parte das empresas tem transformado profundamente a missão da área de tecnologia da informação das empresas, seja através da disponibilização do conteúdo de sistemas legados, processamentos transacional e analítico, construção de novas interfaces.

### **1.2. Contexto da Análise**

Visando definir o contexto da análise deste trabalho é necessário situá-lo na indústria do software, mais particularmente na própria evolução da Web. A Web nasceu em 1991 no laboratório CERN, na Suíça. Seu criador, Tim Berners-Lee, a concebeu apenas como “uma linguagem que serviria para interligar computadores do laboratório e outras instituições de pesquisa e exibir documentos científicos de forma simples e fácil de acessar” [W3C01].

A Web evoluiu rapidamente. Em 1993 já era comum em universidades que estudantes fizessem páginas com informações pessoais. O que determinou seu crescimento foi a criação

de um programa chamado *Mosaic*, que permitia o acesso a Web num ambiente gráfico. Antes do *Mosaic* só era possível exibir textos na Web.

A chave do sucesso da World Wide Web é o hipertexto. Os textos e imagens são interligados através de palavras-chave, tornando a navegação simples e agradável. Os endereços de Web sempre se iniciam com `http://` (*Hipertext Transfer Protocol*). Seu formato mais comum é algo como `http://www. xxxxxx.com.br`, onde:

- `www`: (*World Wide Web*) indica que o endereço pertence a Web (não obrigatório);
- `xxxxxx`: nome da empresa ou instituição que mantém o serviço;
- `com`: indica que é comercial;
- `br`: indica que o endereço é no Brasil.

Novos detalhes e implementações são constantemente divulgados pelo World Wide Web Consortium (W3C) que apresenta as últimas tecnologias de interoperabilidade disponibilizadas [W3C01].

Adicionalmente, o emprego do protocolo TCP/IP que possui um alto grau de interoperabilidade, possibilitou a difusão da web baseada nas seguintes premissas:

- **Independência dos Nós da Rede** - as redes a serem construídas não poderiam parar, mesmo que alguns dos computadores desta rede fossem destruídos;
- **Interconectividade/Interoperabilidade** - os computadores não poderiam estar, todos, fisicamente ligados (por motivos de custo); e,
- **Heterogeneidade** - esta rede deveria suportar diversos tipos de computadores.

Conforme Coulouris [COU94] “o sucesso do protocolo TCP/IP é devido à sua independência da infraestrutura de tecnologia de transmissão, habilitando a interoperabilidade entre redes possibilitando a construção de ligações entre redes heterogêneas”. Tal sucesso está expresso no onipresente uso do TCP/IP no mundo Web.

### 1.3. Desafios da informação no âmbito da web

Os desafios advindos do emprego de soluções de informações baseadas na Web impactam diretamente no esforço de modelagem de ferramentas de suporte a navegação, como é proposta deste trabalho. Conforme Baeza-Yates [BAE99], pode-se dividir tais desafios em duas classes: problemas com os dados propriamente ditos e problemas decorrentes da interação do usuário com os processos de recuperação de informações.

Relativo aos problemas com os dados tem-se a seguinte classificação sugerida por Baeza-Yates:

- **Dados Distribuídos:** devido a natureza intrínseca da Web os dados (e em alguns casos, informações) estão espalhados através de diversos computadores e plataformas. Estes computadores estão interconectados sem uma topologia padrão pré-definida, com uma disponibilidade relativa de banda de transmissão através de um amplo conjunto de redes interligadas.
- **Alta porcentagem de volatilidade dos dados:** devido a dinâmica da Internet, novos computadores (e portanto dados) são adicionados e removidos diariamente (estima-se que 40% da Web tenha alteração a cada mês) - [BKA97]. Adicionalmente, há um elevado índice de links “quebrados” derivados da realocação de páginas quando domínios ou nomes de arquivos são alterados ou expurgados.
- **Volume Elevado de Informações:** o crescimento exponencial de novos sites na Web escalam novos requisitos de hardware e software que são difíceis de serem atendidos.
- **Dados Redundantes e Não-Estruturados:** a maioria das pessoas percebe a Web como um conjunto de hiper-textos harmoniosamente distribuídos. Contudo, devido a natureza não-estruturada das páginas HTML, muitos dados são repetidos (copiados ou espelhados) ou apresentam valores muito similares. Estima-se que aproximadamente 30% das páginas Web são redundantes [BKA97].

- **Qualidade dos Dados:** a Web pode ser encarada como uma nova mídia de publicação de notícias. Porém, e este é o principal problema da Web e particularmente das instituições, não há um trabalho de editoração e avaliação da informação. Assim, os dados podem ser falsos, inválidos (muito antigos ou fora de uso), escritos sem refinamento (pouca clareza) ou, tipicamente com muitos erros de diferentes origens (erros gramaticais, de tradução, de reconhecimento ótico – OCR).
- **Heterogeneidade dos Dados** – além das questões anteriores, a globalização da Web tem apontado para a necessidade de operar com informações em diversas mídias e em diversos padrões. Além disso, todas estas dificuldades são incrementadas pela diferença entre as diversas línguas com a diversidade de alfabetos (Ex. Chinês).

Muitos dos problemas citados (tais como a variedade dos tipos de dados e baixa qualidade das informações) não são solúveis simplesmente pela melhoria de softwares. De fato, muitos destes problemas não deverão mudar no futuro porque são problemas inerentes a própria diversidade de comportamento humano.

A segunda classe de problemas consiste daqueles derivados da interação do usuário com as ferramentas de suporte à busca de informações e podem ser divididos basicamente em dois tipos:

- a) Como definir uma consulta, e;
- b) Como interpretar a resposta proporcionada pelas ferramentas de suporte.

Considerando a diversidade apresentada pela semântica de um conteúdo de um documento Web não é fácil precisar uma consulta estruturada para a informação desejada. Para piorar, mesmo que o usuário consiga estruturar uma consulta conveniente para a informação requerida, irá deparar-se com um número elevado de páginas “candidatas a resposta”.

A dificuldade subsequente será como classificar estas páginas, quais páginas devem ser filtradas pela sua irrelevância, quais páginas devem ser efetivamente acessadas pelo usuário? Tanto no âmbito da Web quanto no âmbito das instituições, esta questão é significativa e tem poucas propostas de solução.

#### 1.4. Foco de Avaliação

O foco de avaliação deste trabalho são os Portais Corporativos e a possibilidade de apoiar a navegação dos usuários a partir da análise dos registros de log do servidor Web. Os Portais Corporativos são apontados como um instrumento de integração das organizações disponibilizando informações, serviços e aplicações a partir das Intranets que atualmente estão proliferando nas empresas de médio e grande porte. Para caracterizar claramente este foco de avaliação deve-se, primeiramente, conceituar claramente os conceitos Internet e Intranet, além da definição correlata de Extranet.

A Internet vem se constituindo no grande fenômeno deste final de século, pela possibilidade de conectar, virtualmente, pessoas ou organizações em qualquer ponto do mundo, de forma simples, padronizada, e com baixo custo a todos os envolvidos. O uso comercial da Internet evoluiu bastante desde as primeiras e tímidas páginas de divulgação institucional, e mostra sinais de ter encontrado o caminho e o ritmo certos. Neste contexto, os conceitos de Internet, Intranet e Extranet podem ser formalizados, conforme segue:

- **Internet** - As informações ficam disponíveis a qualquer visitante que acessa o site, permitindo a consulta e atualização direta às bases de dados;
- **Intranet** - Rede que funciona como a Internet, só que voltada para dentro de uma empresa;
- **Extranet** - É a extensão da Intranet para fora da empresa de forma segura e controlada.

Refinando estes conceitos de Internet, Intranet e Extranet, pode-se montar uma tabela comparativa, apresentada a seguir. Denota-se, então, claramente as diferenças entre os três conceitos. Considerando seu crescimento e disseminação, percebe-se que levou menos que 5 anos para a Internet alcançar 50 milhões de usuários. Ao compararmos com o crescimento de outras tecnologias percebemos a relevância da Internet para empresas e nações:

- TV a cabo: levou 10 anos para atingir 50 milhões de usuários
- TV tradicional: 13 anos
- Telefone: 48 anos

Conceito	Abrangência	Descrição
Internet	Todos os usuários Web	As informações ficam disponíveis a qualquer visitante que acessa o site, permitindo a consulta e atualização direta às bases de dados. É o modelo ideal para as vendas diretas ao consumidor final, serviços de suporte técnico e de atendimento ao cliente.
Intranet	Funcionários de uma empresa	Rede que funciona como a Internet, só que voltada para dentro de uma empresa. A Intranet deve otimizar os processos de administração e troca de informações permitindo que se compartilhe informações internas usando qualquer <i>browser</i> e reduzindo drasticamente o gasto com telecomunicações entre filiais. Outra vantagem da Intranet é o fato de que as informações que nela circulam não estão disponíveis à qualquer um com acesso à Internet.
Extranet	Funcionários e Usuários autorizados a acessar uma Intranet remotamente	É a extensão da Intranet para fora da empresa de forma segura e controlada. A sua empresa poderá usar a Extranet para prover informações privadas à seus parceiros ou consumidores sem ter que usar o telefone, poupando assim tempo e dinheiro. É através da extranet, por exemplo, que as empresas disponibilizam dados para seus fornecedores e parceiros comerciais.

*Tabela 1.1 – Conceitos Internet, Intranet e Extranet*

Não há dúvidas que o conteúdo da Internet atualmente excede o de qualquer biblioteca do mundo em informação<sup>1</sup>, e se tornará rapidamente nos próximos anos a principal fonte de conhecimento usada pela mente humana.

### 1.5. Organização

Este trabalho foi estruturado de forma a avançar no contexto da tecnologia da informação situando historicamente o surgimento do conceito dos Portais Corporativos, bem como agregando a atual necessidade de busca de conhecimento com eficiência e rapidez.

Neste capítulo foi apresentado o paradigma da conectividade representado pela Web, um meio de comunicação mundialmente abrangente, relativamente seguro, crescentemente interativo e de baixo custo, situando o contexto da análise deste trabalho. Na sequência apresentam-se os principais desafios advindos do emprego de soluções de informações baseadas na Web, notadamente as questões relativas a distribuição dos dados, o alto percentual de volatilidade, volume elevado de informações, redundância, qualidade e heterogeneidade dos dados.

Tais desafios apontam para um problema comum de identificação, de preferência rapidamente, da informação necessária. Por fim, buscando apontar uma alternativa computacional à esse problema denota-se o foco de avaliação deste trabalho no contexto dos Portais Corporativos e na possibilidade de apoiar a navegação dos usuários a partir da análise dos registros de log dos servidores Web. O próximo capítulo irá explorar em maior profundidade o contexto da Web nas corporações buscando identificar suas principais características e propriedades.

O capítulo 2 foca o trabalho no contexto das corporações, caracterizando o conceito de um Portal Corporativo e apontando as alternativas de abordagem, sejam elas padrões de mercado (*standard*) ou soluções desenvolvidas internamente. Também são apresentadas as características de um Portal Corporativo e o detalhamento de um caso especial de Portal Corporativo o *Business-to-Employee* (B2E). A partir deste contexto é definida a abrangência

---

<sup>1</sup> Estimativas apontavam mais de 350 milhões de páginas – Jul/98 [BAE99]

da Gestão do Conhecimento demonstrando a necessidade dos usuários de Portais Corporativos de encontrar as informações com agilidade e eficiência.

O capítulo 3 – Portal Corporativo – Esquema Conceitual – traz um estudo das diversas visões que a tecnologia de portais apresenta atualmente buscando focar em alguns padrões e requisitos que possam facilitar a identificação do escopo de atuação. São abordadas as características inerentes aos Portais Públicos e Privados, privilegiando este último. O Portal Corporativo é tipificado de forma a identificar com maior propriedade suas funções principais consolidando os diversos conceitos disponíveis.

O capítulo seguinte – Análise de Soluções – faz uma avaliação das diferentes abordagens para o suporte à navegação dos usuários num site de uma forma geral e, notadamente, num Portal Corporativo, detalhando as características de customização de sites encontradas atualmente. São avaliadas alternativas de customização definidas na indústria de informática bem como apresentadas algumas das principais ferramentas de mercado. Demonstra-se com estas informações que as abordagens focam em *Ferramentas Baseadas em Regras e Ferramentas baseadas em Filtragem Colaborativa*. Percebe-se o potencial de emprego de uma ferramenta de suporte à navegação baseada no perfil do usuário foco deste trabalho.

O capítulo 5 aborda a proposta de solução propriamente dita, apresentando a arquitetura planejada a partir dos requisitos a serem contemplados bem como uma justificativa dos requisitos não atendidos. Apresenta-se o funcionamento da ferramenta que recebe como entrada a log do servidor web e organiza seus dados de forma que seja possível extrair informações como o perfil de determinado grupo de usuários. A proposta de solução parte da premissa de que é possível modelar o site como um grafo dirigido, com as páginas correspondendo aos vértices deste grafo e as arestas aos links. Este capítulo detalha a implementação computacional mencionando os diversos componentes envolvidos bem como demonstrando o processo utilizado. Por fim são avaliados os resultados obtidos com considerações a cerca da utilização desta solução.

Ao final é apresentada uma conclusão detalhada deste trabalho contemplando o principal resultado, caracterizado pela proposta de ferramenta de suporte à navegação baseada

no perfil do, bem como abordando alguns resultados indiretos tais como uma sugestão de tipificação de portais. Também são indicadas algumas pesquisas complementares que podem ser efetivadas futuramente continuando com esta pesquisa.

## Capítulo 2

# WEB NO CONTEXTO DAS CORPORAÇÕES

Este capítulo apresenta a web no contexto do ambiente das organizações, caracterizando o conceito de um Portal Corporativo e apontando as variantes de solução, sejam elas padrões de mercado ou soluções desenvolvidas internamente.

Também são apresentadas as características de um Portal Corporativo e o detalhamento de um caso especial de Portal Corporativo o *Business-to-Employee* (B2E).

### 2.1. A Web no Ambiente Corporativo

A web no ambiente corporativo – caracterizada como Intranet conforme visto no capítulo anterior - integra aplicações internas como e-mail, acesso a banco de dados e gestão de documentos, com aplicações externas como serviços de notícias e aplicações internas acessadas através deste novo canal. É uma interface baseada na Web que permite ao usuário acessar todas estas aplicações por meio de um navegador (*browser*) no computador pessoal.

Nesta interface, um usuário verifica as normas relevantes e os produtos que tiveram alterações no processo de comercialização, pesquisa a cotação das ações da empresa, checa a sala de reuniões disponível e recebe um pedido de avaliação de desenvolvimento profissional - tudo isso através do *browser* em seu *desktop*. É uma nova geração de Intranet, também conhecida como **Portal Corporativo**. Pode-se afirmar que o *browser* se torna um painel de instrumentos para as tarefas diárias.

A função de um Portal Corporativo é centralizar as informações importantes no desktop do funcionário. Ele também deve filtrar a maioria das informações irrelevantes, como os milhares de páginas de Intranet ou os e-mails errantes que se destinam a funcionários de outros departamentos.

É muito difícil criar um portal que execute estas tarefas em toda a sua plenitude. Segundo Hadley Reynolds, analista do Delphi Group, os Portais Corporativos requerem oito camadas de tecnologia [REY00]. Wayne Eckerson, analista do Patricia Seybold Group, diz que os fornecedores têm que poder disponibilizar 15 áreas de funcionalidade para que suas soluções sejam consideradas suficientes [ECK01]. No item 2.3. Características do Portal Corporativo, são apresentadas tanto as camadas propostas por Reynolds quanto as funcionalidades abordadas por Eckerson.

São as camadas e as funções que tornam um portal tão complexo. Não se trata apenas de uma página Web com um punhado de links para sites interessantes. É uma página que dá aos usuários dados de fontes estruturadas como aplicações de bancos de dados, planilhas e informações derivadas de ferramentas tais como *Enterprise Resource Planning (ERP)*, assim como dados não estruturados de documentos, e-mail e páginas Web. Inclui uma ferramenta de pesquisa e muitas listagens de milhares de páginas de informações baseadas na Intranet, divididas em categorias e classificadas conforme um padrão corporativo. Pode incluir também um alimentador de notícias (*news feeder*) do mundo externo e produtos *groupware* que permitam aos profissionais colaborarem em projetos. A idéia de ter tanta coisa acontecendo em uma janela constitui um novo modelo de computação *desktop*, conforme exposto por Reynolds [REY00].

## **2.2. Soluções *standard* x Soluções caseiras**

Emboras as opiniões dos analistas variem em termos da proporção do trabalho que os fornecedores podem realizar para uma empresa, uma solução fácil está há anos de distância, afirma Reynolds [REY00]. "Não existe um pacote de Portal Corporativo". Muitos fornecedores, porém, estão oferecendo respostas parciais, baseados em sua competência e seu conhecimento a cerca do mercado. "As versões iniciais da maioria dos produtos de portal são versões ligeiramente reformuladas de ferramentas de '*business intelligence*' ou 'gestão de documentos'", escreveu Eckerson - [ECK01].

A arquiteta de sistemas Anne Jean-Blanc conta que a pesquisa para a montagem de um Portal Corporativo na Caterpillar envolveu uma análise de produtos das empresas Verity, Dataware Technologies, Viador e Plumtree Software. A instalação de um Portal Corporativo pode custar caro. Com 65.000 funcionários, a Caterpillar gastaria facilmente milhões de dólares em licenças de software que custam US\$ 10 ou US\$ 20 por usuário [PLU00]. Criar e implantar um bom portal não é uma tarefa simples e as empresas que tentam criar portais por conta própria podem levar anos para terminar o trabalho, a um elevado custo. Este desafio é particularmente encarado pelas administrações contemporâneas que buscam fortemente a centralização das operações, objetivando redução de custos.

Construir um Portal deve ser tarefa para uma equipe especializada que agregue:

- **Expertise Técnico** - caracterizado pelo conhecimento das diversas ferramentas disponíveis na Web;
- **Visão Organizacional** - contemplando um profundo conhecimento da instituição e do seu *modus operandi*;
- **Formalismo Taxonômico** – busca constante da formalização para os tópicos e itens dispostos no Portal de forma a convergir conceitos e evitar a disparidade de estruturação;
- **Inovação** – elaborar soluções simples que possam suportar o trabalho do empregado a partir de releituras do processo atual

A partir destes requisitos, algumas funcionalidades específicas podem ser exploradas, em especial a possibilidade de apresentar informações de forma organizada e compreensível.

### 2.3. Características do Portal Corporativo

Os Portais Corporativos devem ser estruturados em várias camadas de múltiplas tecnologias, o que dificulta sua criação. Para melhor compreender este universo, segue

quadro contemplando as possíveis características de um Portal conforme proposto por Hadley Reynolds do The Delphi Group Inc., Boston [REY00].

<b>Característica</b>	<b>Descrição</b>
<b>Apresentação</b>	Tecnologias padrões de exibição na Web tais como HTML, JavaScript e applets ou Cascading Style Sheets. Também são consideradas ferramentas de apresentação de e-mail ( <i>viewer</i> ) e de visualização de gráficos originados em processos OLAP destinados a Web.
<b>Personalização</b>	Agentes que filtram informações para usuários individuais. Eles podem sugerir os focos de interesse dos usuários e aprender a partir do que os usuários fazem.
<b>Colaboração</b>	Tecnologias de <i>groupware</i> como discussões, sessões de <i>chat</i> e bibliotecas de projetos.
<b>Processo</b>	Tecnologias como transações online que são os <i>engines</i> de diversos processos de negócios ou <i>workflow</i> .
<b>Publicação &amp; Distribuição</b>	Depósitos de documentos em formatos portáteis como <i>Portable Document Format</i> e dispositivos de publicação/assinatura ou outras maneiras de "publicar" informações.
<b>Busca</b>	Ferramentas de busca baseada no texto completo ( <i>full-text</i> ) e as que pesquisam descrições de documentos e outro tipo de conteúdo.
<b>Categorização</b>	Ferramentas para criar e manter categorias. As categorias têm que ser variadas para audiências diferentes que examinam os mesmos documentos e dados de maneiras diferentes.
<b>Integração</b>	Ferramentas para acessar dados de diversas fontes tais como bancos de dados relacionais e pacotes ERP. Além disso, ferramentas que disponibilizam dados externos ( <i>data feeds</i> ), como notícias ou cotações de ações. Integra dados estruturados e não estruturados de sistemas de arquivos, Web servers e e-mail.

*Tabela 2.1 – Características de um Portal Corporativo*

As camadas agrupam algumas tecnologias conforme sugestão de Reynolds. Busca-se, neste caso, identificar as principais características de um Portal como Personalização, Colaboração, Publicação e Distribuição dentre outras. Sob outra abordagem, conforme proposto por Eckerson [ECK01], é possível identificar 15 funcionalidades para uma solução de Portal Corporativo, detalhadas no item 3.4. Requisitos mínimos de um Portal Corporativo.

#### **2.4. Portal *Business-to-Employee* (B2E)**

Uma derivação no conceito de Portal Corporativo é atualmente encontrada na literatura como sendo o Portal *Business-to-Employee* ou B2E. Este conceito tem sido disseminado pelo advento da Nova Economia ou economia digital (*e-economy*) e por seus adeptos.

Conceitos como *business-to-business* (B2B) ou *business-to-consumer* (B2C) converteram-se no dia a dia da atividade empresarial transformando as relações entre as companhias e os clientes. Com o objetivo de revolucionar a interação entre organizações e empregados, surge no cenário empresarial um conceito que foi classificado por muitos como o novo estágio da revolução da Intranet: o *Business-to-Employee* (B2E).

Otimizar ao máximo a eficiência de operação e a produtividade de um trabalhador ao mesmo tempo em que se reduz a complexidade de suas tarefas diárias é a finalidade máxima dos portais B2E. Um conceito que supõe caracterizar-se como a porta de entrada para a gestão dos Recursos Humanos das corporações na revolução dos ditos “canais eletrônicos” ou *e-solutions*. Os portais B2E possibilitam migrar de uma administração tradicional para um novo conceito guiada pelas diretrizes da Nova Economia.

Os funcionários são o ativo mais importante de uma empresa e melhorar sua relação com a mesma é um elemento crítico para incrementar a eficiência. Nos primeiros momentos as empresas implementam apenas Intranets e Extranets corporativas. A este respeito o Gartner Group estima que 80% das maiores empresas do mundo já contam com um Portal Corporativo com capacidade de disponibilização facilitada de informações para os funcionários. Segundo dados apresentados pela Meryll Lynch, as companhias investiram mais no desenvolvimento de seus portais do que em sistemas de ERP [WHI01].

Os Portais Corporativos passam por um processo de aprimoramento que está culminando no atual estágio de evolução caracterizado pela real capacidade de acesso a múltiplas informações, de maneira personalizada e em tempo real. Comparativamente à evolução gerada pelo auto-atendimento para o mercado financeiro, a autogestão de informações representa este marco no processo de implantação de Portais Corporativos.

Um recente estudo realizado pelo *Butler Group* [WHI01] ouviu companhias do Reino Unido e revelou que 70% delas já têm um portal corporativo ou projetos para construí-lo no período de um ano. Apenas 6% não têm sequer planos de criar um portal. O número de usuários por portal variou entre 10 mil e 22 mil por companhia, sendo que algumas das empresas querem atingir até 90 mil usuários.

Ainda, segundo o *Butler Group*, os portais internos informativos, além de esclarecer as dúvidas dos funcionários da empresa, também são capazes de incentivar o trabalho colaborativo, bem como suportar o gerenciamento do conteúdo e das informações envolvendo a companhia.

Além disso, este estudo apontou as características essenciais em um Portal Corporativo: gestão de conhecimento, segurança intrínseca ao âmbito da empresa e acesso centralizado a informações e aplicativos. Deve-se registrar, neste ponto, que o foco de aplicação deste trabalho será dirigido para a gestão de conhecimento desenvolvida em instituições financeiras.

## **2.5. Gestão do Conhecimento**

A gestão de conhecimento deve ser utilizada pelas organizações para organizar, codificar e disseminar o saber armazenado nos documentos que representa a experiência e o conhecimento acumulados nas pessoas. Uma definição mais formal poderá ser a que é avançada por Carl Frappaolo (*Delphi Group*): Gestão do Conhecimento é um conjunto de ferramentas para a automatização das relações dedutivas ou inerentes entre os usuários, os processos e as páginas de informação - [FRA01].

A gestão de conhecimento apresenta-se como uma grande promessa para a indústria em geral e para o setor financeiro em particular. Focando o estudo especificamente neste

setor, percebe-se também um grande desafio para uma indústria, que depende em grande parte de produtos intangíveis e serviços. Estes produtos e serviços, por sua vez, derivam da capacidade de uma organização agregar capital intelectual e promover inovação.

Numa avaliação realizada pela revista *Knowledge Management* acerca do progresso da gestão do conhecimento aplicada ao setor financeiro – [Set/2000] - era aparente que a indústria em geral estava atrasada comparativamente ao passo marcado pelas organizações em outros setores econômicos. Como Robert Taylor, diretor de soluções da Unisys, disse: "Poucas instituições financeiras lançaram programas sérios de gestão do conhecimento. Das que o fizeram, a maioria está nos primeiros estágios, e o seu enfoque tende a ser de infraestrutura e não estratégico" [KDD01]. Exatamente um ano depois e baseado numa avaliação detalhada sobre a gestão do conhecimento no setor financeiro efetuada pela *Chartered Institute of Bankers (CIB) and Xerox ("Winning through knowledge" na Financial World, Março 2001)* pode-se identificar diversos exemplos práticos.

Conforme artigo de David J. Skyrme, um trabalho pioneiro na gestão do conhecimento em Portais Corporativos é o da empresa *Skandia*, uma empresa de seguros de âmbito mundial. "O *Intellectual Capital Navigator* da *Skandia* foi a ferramenta pioneira para identificar, gerir e medir o mundo dos bens intangíveis e geradores de valor de uma organização" – [SKI01].

É difícil identificar um produto líder claro no campo da gestão de conhecimento, já que o setor financeiro é composto por numerosas entidades operando em negócios que vão desde os ciber-bancos (bancos totalmente orientados ao canal eletrônico – *cyberspace*), passando por empresas de gestão de investimento institucional até os grandes bancos de varejo, todas com programas e prioridades diversas no que diz respeito à gestão de conhecimento. Contudo é também aparente que, à parte destas poucas notáveis exceções, a maioria das instituições financeiras ainda têm um longo caminho a percorrer nos seus esforços de gestão de conhecimento.

O setor financeiro é uma indústria dependente da tecnologia da informação (*IT-Information Technology*) e pessoas, por isso há imenso potencial para a gestão do conhecimento. Contudo, há uma opinião forte das diversas consultorias que a adoção da gestão do conhecimento tem sido marcadamente mais lenta no setor financeiro do que em

muitas outras indústrias, conforme afirmado por James Digges, consultor executivo da KPMG Consulting [DIG01].

Também parece haver uma confusão sobre o que a gestão do conhecimento realmente significa no setor financeiro, como afirma Ibrahim Gogus, diretor de Gestão de Conhecimento na Oracle, "Por quê arriscar mudar processos de negócio, estruturas organizacionais e sistemas de *legacy* bem estabelecidos, que funcionaram bem durante décadas, sem especificar benefícios tangíveis?" – [GOG01].

Mesmo assim a promessa da gestão de conhecimento permanece tão viva para as instituições financeiras como para as organizações que operam em outras indústrias. Vantagens como redução de custos (através da melhoria de processos e da realocação de pessoal; criação de valor agregado (reduzindo a quantidade de tempo necessário para os empregados reaprenderem lições documentadas) e aumento de receitas (na medida em que o conhecimento da empresa pode ser facilmente empacotado e vendido), continuam constituindo como os potenciais benefícios a serem almejados.

Talvez a relutância aparente do setor financeiro em adotar a gestão de conhecimento como estratégia prioritária na organização possa ser explicada por alguns problemas específicos que devem ser ultrapassados. A gestão de conhecimento ainda deve conquistar os executivos seniores quanto ao efetivo retorno do capital investido (*roi – return of investment*), bem como superar a dificuldade representada pela quebra das tradicionais barreiras culturais inerentes à partilha do conhecimento que ainda existem na indústria.

O nível de fragmentação atual no que diz respeito a isso torna a venda cruzada extremamente difícil ao mesmo tempo que o conflito cultural e organizacional causado por processos de fusão ou *take-over* apenas agravam o problema. Por fim, organizações no setor financeiro também têm de lidar com os problemas derivados do relacionamento de diversos profissionais geograficamente dispersos, bem como numerosas restrições relacionadas com a troca de conhecimento impostas por regulamentações e questões de confidencialidade.

Para que a gestão de conhecimento tenha um verdadeiro impacto, o ônus deve ser da alta administração que deve iniciar um processo de mudança da cultura organizacional bem

programada. Esta atitude tem de tornar claros os benefícios de uma organização que compartilha conhecimento:

- Incremento da vantagem competitiva;
- Aumenta a eficácia do controle de custos;
- Crescimento pessoal; e,
- Critério para avaliação de desempenho.

Uma evolução deste conceito, onde a gestão de conhecimento cria raízes na organização, é determinada pela empresa que luta para criar um ambiente rico em conhecimento, que ofereça oportunidades únicas para o crescimento pessoal e de lealdade. Dificilmente haverá um lugar melhor para trabalhar do que um onde o conhecimento das pessoas seja desejado e necessário, e da mesma forma, onde seja possível aprender baseado na experiência e no conhecimento de outros – fato corroborado pela implementação de processos colaborativos (*groupware*).

Como sempre, a tecnologia tem um papel crítico a desempenhar. O uso de taxonomias e uma boa arquitetura de informação são cruciais, enquanto o conteúdo é a chave para atrair e reter usuários de um Portal Corporativo que empregue os conceitos de gestão do conhecimento.

O trabalho do conhecimento é geralmente não estruturado, imprevisível, *ad hoc*, colaborativo e contextual. Mais do que nunca haverá necessidade de formação de profissionais capacitados na combinação de informações, interpretação dos valores, definição de conclusões e tomadas de decisão baseadas em informações de múltiplas fontes.

Talvez o maior perigo no progresso da gestão de conhecimento no setor financeiro seja a continuada confusão sobre o que realmente significa. Historicamente, o setor financeiro parece focar na tecnologia em vez de focar nas intangíveis e culturais implicações da gestão do conhecimento. Uma recente avaliação da IDC estimava que os gastos em sistemas de gestão de conhecimento no setor bancário Europeu iria crescer de \$155.4 milhões este ano para \$511.4 milhões em 2004. O IDC estima, ainda, que a indústria de gestão do conhecimento irá atingir cifras da ordem de 6 bilhões de dólares em 2002 - [STE02].

Embora estas descobertas não sugiram por si só que as instituições financeiras estão evoluindo para um melhor entendimento sobre a gestão de conhecimento ainda assim parece que o interesse e o empenho na gestão de conhecimento está crescendo no setor financeiro.

## 2.6. Conclusão

Neste capítulo foi apresentado como a web pode ser vista no contexto das corporações caracterizando o conceito de Portal Corporativo. Foram abordadas as características e funcionalidades de um Portal bem como o conceito derivado da nova economia de *Business-to-employee* (B2E). Por fim, foi enfatizado a necessidade de foco dos Portais Corporativos na gestão do conhecimento visando garantir um diferencial de mercado para as demais instituições.

No próximo item será abordado o esquema conceitual de um Portal Corporativo abordando quesitos tais como sua definição, os tipos relacionados ao contexto de sua utilização (público ou corporativo), suas classes e sub-classes e, por fim, os requisitos mínimos de um Portal.

## Capítulo 3

# PORTAL CORPORATIVO – ESQUEMA CONCEITUAL

Com os inúmeros avanços tecnológicos, muitas instituições sofrem com o excesso de informações, sendo imprescindível a aplicação da gestão do conhecimento para administrar esse caos informacional do mundo digital. Muitas vezes as informações estão armazenadas em equipamentos de informática de forma não integrada, espalhadas em seus bancos de dados, dificultando seu acesso e, conseqüentemente, o desempenho das atividades necessárias ao pleno funcionamento da instituição.

Devido à proliferação de arquivos eletrônicos produzidos individualmente pelos funcionários, sem o intuito de compartilhamento de informações, e de várias aplicações de bancos de dados, desenvolvidas ao longo do tempo para atender a demandas específicas, falta, a essas instituições, uma visão global de seus próprios dados e informações.

Recentemente, surgiu um novo conceito, o Portal Corporativo, que integra dados não estruturados aos dados estruturados dos bancos de dados institucionais, fornecendo acesso às informações a partir de uma interface individualizada, disponível na rede hipertextual corporativa - Intranet.

O Portal Corporativo é tido, por Collins, como o mais importante projeto de gestão da informação da próxima década [COL99]. Como uma evolução do *data warehouse*, o portal corporativo estende sua aplicação à Intranet e se constitui em um único ponto de acesso a todos os recursos de informação e conhecimento em uma instituição.

Embora seja uma tecnologia muito recente, vários são os benefícios, apontados por fornecedores e consultores de informática, associados aos portais corporativos. Dentre esses benefícios, destaca-se a facilidade de acesso às informações distribuídas nos diversos sistemas, arquivos e bases de dados institucionais.

Para conseguir concretizar esse benefício, é fundamental que o projeto do portal corporativo leve em consideração a interação dos usuários com sua interface. Sua capacidade de facilitar o acesso dos usuários às informações institucionais está intrinsecamente relacionada à facilidade de uso, aprendizado e satisfação do usuário, isto é, à usabilidade de sua interface web.

O termo usabilidade foi definido na norma ISO/IEC 9126, sobre qualidade de software, como "um conjunto de atributos de software relacionado ao esforço necessário para seu uso e para o julgamento individual de tal uso por determinado conjunto de usuários." [ISO91]. A partir de então, esse termo passou a fazer parte do vocabulário técnico de várias áreas do conhecimento, tais como tecnologia da informação, ergonomia, interação homem-computador e psicologia aplicada, tendo sido traduzido literalmente para diversos idiomas.

O conceito de usabilidade evoluiu e foi redefinido na norma ISO 9241-11 *Guidelines on Usability* como "a capacidade de um produto ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso" [ISO98].

Como ferramentas auxiliares de projeto de sistemas centrados no usuário, existem vários métodos de avaliação de usabilidade, cujos objetivos são verificar a eficiência e a eficácia da interação usuário-computador durante a realização de tarefas; e identificar, de forma subjetiva, o grau de satisfação dos usuários com o sistema.

A variedade de métodos de avaliação de usabilidade existentes, sua necessidade de adaptação ao contexto de uso e a evolução tecnológica da gestão da informação, representada

pelo portal corporativo, deram origem a um problema de pesquisa, relacionado à identificação e adaptação de métodos de avaliação de usabilidade para o contexto de portais corporativos.

### 3.1. Definição de Portal Corporativo

Há três ou quatro anos, o que hoje é chamado de portal era conhecido como máquina de busca (*search engine*), cujo objetivo era facilitar o acesso às informações contidas em documentos espalhados pela Internet. Inicialmente, as máquinas de busca possibilitavam ao usuário da Internet localizar documentos a partir de pesquisas booleanas e navegação associativa entre links.

Para reduzir ainda mais o tempo de busca na Internet e auxiliar os usuários menos experientes, vários sites de busca incluíram categorias, isto é, passaram a filtrar sites e documentos em grupos pré-configurados de acordo com seu conteúdo - esportes, meteorologia, turismo, finanças, notícias, cultura, dentre outros.

O passo seguinte foi a integração de outras funções, como, por exemplo, as comunidades virtuais e suas listas de discussão, chats em tempo real, possibilidade de personalização dos sites de busca (Exemplo: My Yahoo) e acesso a conteúdos especializados e comerciais.

Essa nova concepção mais abrangente é que passou a ser chamada de portal. Reynolds & Koulopoulos [REY00] identificam as seguintes fases do progresso do portal web:

- pesquisa booleana,
- navegação por categorias,
- personalização e,
- funções expandidas para outras áreas dos mundos informacionais e comerciais (como por exemplo a integração com sistemas de gerenciamento de produção – *supply chain*).

Essa evolução dos portais web, ou portais públicos, chamou a atenção da comunidade corporativa, a qual vislumbrou a possibilidade de utilização dessa mesma tecnologia para

organizar e facilitar o acesso às informações internas da empresa. Por ser um conceito muito recente, a terminologia relacionada com os portais corporativos ainda não se estabilizou.

Os conceitos **portal corporativo**, **portal de informações corporativas**, **portal de negócios** e **portal de informações empresariais** são utilizados na literatura, algumas vezes, como sinônimos. Segundo Firestone, “o processo de definição do portal corporativo, como qualquer outra estratégia de negócios, é um processo político, isto é, uma tentativa de persuadir os usuários e os investidores da área de tecnologia da informação que uma definição é mais adequada do que outra, favorecendo os interesses de um ou outro fornecedor/consultor” [EIP00].

A aceitação geral de uma definição feita por determinado fornecedor pode levar o público a concluir que o produto da concorrência, por deixar de ter esta ou aquela característica, não é realmente um portal corporativo, por exemplo. Deixando de lado o aspecto político entre concorrentes e enfatizando as características inerentes de cada termo, a seguir são apresentadas as definições consideradas mais relevantes para este trabalho.

O conceito **Portal de Informações Empresariais** foi utilizado pela primeira vez em um relatório da empresa de consultoria Merrill Lynch, elaborado por Shilakes & Tylman em novembro de 1998, onde consta a seguinte definição: “Portais de Informações Empresariais são aplicativos que permitem às empresas liberar informações armazenadas interna e externamente, provendo aos usuários uma única via de acesso à informação personalizada necessária para a tomada de decisões de negócios” [SHI99].

Considera-se neste trabalho que o Portal de Informações Empresariais ou EIP (*Enterprise Information Portal*) é uma oportunidade emergente de mercado, uma estratégia de negócios, constituindo-se em um conjunto de aplicativos de software que consolida, gerencia, analisa e distribui informações não só internamente, como também para o ambiente externo à organização, incluindo ferramentas de inteligência de negócios, gestão de conteúdo, data warehouse, gestão de dados e informações [SHI99]. Apesar dessa definição ser abrangente o suficiente para aliar duas funções dos portais corporativos, suporte à decisão e processamento cooperativo, o relatório, como um todo, não dá ênfase suficiente ao aspecto cooperativo, apesar de mencioná-lo.

Nessa mesma linha de raciocínio, White define o EIP como uma ferramenta que provê, aos usuários de negócios, uma única interface web às informações corporativas espalhadas pela empresa [DAI99]. Dentro desse conceito mais genérico, White ressalta as duas funções mencionadas anteriormente, subdividindo os EIPs em duas categorias: EIP para processamento de decisões e EIP para processamento cooperativo [DAI99].

Para White, o EIP para processamento de decisões auxilia executivos, gerentes e analistas de negócios no acesso às informações necessárias para a tomada de decisões de negócios, enquanto o EIP para processamento cooperativo organiza e compartilha informações de grupos de trabalho, tais como mensagens de correio eletrônico, relatórios, memorandos, atas de reunião, dentre outros [DAI99].

Eckerson, por sua vez, emprega outro termo - Portal de Negócios - e o define como um aplicativo capaz de proporcionar aos usuários um único ponto de acesso a qualquer informação necessária aos negócios, esteja ela dentro ou fora da corporação. Eckerson compara o Portal de Negócios a um “shopping center para trabalhadores do conhecimento” explicando que a maioria dos consumidores prefere fazer suas compras em um shopping, por ter certeza de lá poder encontrar tudo de que precisa, evitando a busca cansativa em sucessivas lojas em diferentes localidades [PSG00]. Sua concepção difere do ponto de vista de White, Shilakes & Tylman por dar pouca ou quase nenhuma ênfase ao aspecto colaborativo, limitando o conceito de portal como uma via de acesso às informações estruturadas e não estruturadas, por meio de uma interface web [PSG00].

Uma linha de pensamento um pouco diferente, defendida por Murray [MUR99], considera o portal corporativo mais do que uma via de acesso às informações corporativas, como definem Shilakes & Tylman [SHI99]. Murray afirma que os portais voltados apenas para o conteúdo são inadequados ao mercado corporativo e que “os portais corporativos devem nos conectar não apenas a tudo de que necessitamos, mas a todos que necessitamos, e proporcionar todas as ferramentas necessárias para que possamos trabalhar juntos” [MUR99].

Murray identifica quatro tipos distintos de portais empresariais: portais de informações, portais cooperativos, portais de especialistas e portais do conhecimento

[MUR99]. Os portais de informações provêm acesso à informação; os portais cooperativos fornecem ferramentas de processamento cooperativo; os portais de especialistas conectam pessoas, com base em suas experiências e interesses; finalmente, os portais do conhecimento combinam todas as características dos anteriores [MUR99].

Com essa subdivisão, Murray parece estar mais interessado em um portal capaz de atender a todas as expectativas funcionais dos usuários corporativos, e não apenas em uma ferramenta de tomada de decisão ou de acesso a informações. Assim como Murray destaca o caráter cooperativo do portal, Reynolds & Koulopoulos vislumbram o portal como um sistema de informações centrado no usuário, integrando e divulgando conhecimentos e experiências de indivíduos e equipes, atendendo, assim, aos padrões atuais de instituições baseadas no conhecimento [REY00]. Para esses autores, o portal corporativo é capaz de aliar o conhecimento explícito contido em arquivos, bases de dados, correspondências, páginas web e aplicativos empresariais ao conhecimento tácito dos times de projeto, das heurísticas profissionais e das comunidades institucionais [REY00].

Da mesma forma que acontece com os consultores e analistas de mercado na área de informática, também os fabricantes e fornecedores de software definem os portais corporativos de maneiras diferentes, de acordo com o enfoque de cada produto, destacando características relacionadas ao suporte à decisão e/ou ao processamento cooperativo. Como concorrentes no mercado, cada fabricante realça as qualidades de seu produto em suas definições, conforme segue.

A Viador define os portais de informações empresariais como “aplicativos que possibilitam às companhias fornecer acesso às informações armazenadas interna e externamente e, ainda, oferecerem aos usuários internos e externos uma única janela para as informações personalizadas necessárias para a tomada de decisões de negócios” [VIA00]. Apesar de usar a nomenclatura EIP (*Enterprise Information Portal*), considerada mais genérica, a Viador destaca claramente o suporte à tomada de decisões como a principal função de seu portal.

A Brio, por outro lado, baseando-se nos requisitos considerados pelo *Patricia Seybold Group* como essenciais para os portais corporativos, define seu portal de informações

empresariais como uma ferramenta que provê, aos usuários, amplo acesso à informação, independentemente de onde esta esteja armazenada, e compartilhamento cooperativo de informações entre todos os usuários da organização, independentemente da localidade em que se encontram [BRI00]. O uso da nomenclatura EIP pela Brio é compatível com as definições de White e Shilakes & Tylman.

A Plumtree Software, por sua vez, define seu portal corporativo como um sistema capaz de reunir, em uma única página web personalizada, todas as informações e ferramentas de produtividade relevantes aos usuários corporativos, integrando ainda aplicativos dinâmicos, tais como relatórios on-line, correio eletrônico, agendas e serviços comerciais [PLU00].

A incorporação de aplicativos cooperativos e a capacidade intrínseca do portal de criar relacionamentos entre informações e prover acesso ao conteúdo da rede corporativa e da Internet, por meio de uma visão lógica de informações provenientes de sistemas diversos, dão ao portal da Plumtree Software um caráter mais genérico, isto é, um portal capaz de suportar tanto o processamento cooperativo como o processamento de decisões. A nomenclatura “Portal Corporativo” demonstra a preocupação da Plumtree Software em diferenciar seu produto dos portais públicos, como Yahoo.

Dada a diversidade de nomenclaturas e definições, é conveniente classificar os portais para melhor entender suas características e os enfoques dados por cada fonte citada.

### **3.2. Portais Públicos ou Corporativos**

Há duas formas de classificar os portais: uma em relação ao contexto de sua utilização (público ou corporativo) e outra em relação às suas funções (suporte à decisão e/ou processamento cooperativo). Como o interesse deste trabalho é o portal corporativo, a classificação quanto à função tomará como base o ambiente institucional. Quanto ao contexto, apesar das semelhanças tecnológicas, os portais públicos e os portais corporativos atendem a grupos de usuários diversos e têm propósitos completamente diferentes.

### 3.2.1. Portal Público

O Portal público - também denominado portal Internet, portal web ou portal de consumidores - provê ao usuário uma única interface à imensa rede de servidores que compõem a Internet. Sua função é atrair, para o seu site, o público em geral que navega na Internet. Quanto maior o número de visitantes, maior a probabilidade do estabelecimento de comunidades virtuais que potencialmente comprarão o que os anunciantes daquele site têm para vender.

Assim como a televisão, o rádio e a mídia impressa, o portal público estabelece um relacionamento unidirecional com seus visitantes e constitui-se em uma mídia adicional para o marketing de produtos. Segundo Eckerson, desde meados de 1990, os portais públicos passaram por três gerações diferentes [SPO00], como mostra o quadro de geração visto abaixo.

Geração	Categoria	Características
Primeira	Referencial	Máquina de busca, com catálogo hierárquico de conteúdo da web. Cada entrada do catálogo contém uma descrição do conteúdo e um link.
Segunda	Personalizado	O usuário, por meio de um identificador e uma senha, pode criar uma visão personalizada do conteúdo do portal, conhecida como "MinhaPágina". Essa visão mostra apenas as categorias que interessam a cada usuário. O portal pode avisar ao usuário sempre que um novo conteúdo for adicionado às categorias por ele assinaladas.
Terceira	Interativo	O portal incorpora aplicativos, tais como correio eletrônico, chat, listas de discussão, cotação da bolsa, comércio eletrônico, leilões, permitindo ao usuário interagir com o portal e com seu provedor de conteúdo. Os usuários podem selecionar essas aplicações para suas páginas pessoais.

*Tabela 3.1 – Gerações dos Portais Públicos*

### 3.2.2. Portal Corporativo

No mundo institucional, o Portal Corporativo tem o propósito de expor e fornecer informações específicas de negócio, dentro de determinado contexto, auxiliando os usuários de sistemas informatizados corporativos a encontrar as informações de que precisam para fazer frente aos concorrentes [REY00]. O Portal Corporativo é considerado por Reynolds & Koulopoulos como uma evolução do uso das Intranets, incorporando, a essa tecnologia, novas ferramentas que possibilitam identificação, captura, armazenamento, recuperação e distribuição de grandes quantidades de informações de múltiplas fontes, internas e externas, para os indivíduos e equipes de uma instituição [REY00].

Assim como os portais públicos, os portais corporativos também passaram por estágios evolutivos, porém em um espaço de tempo bem menor. Eckerson identifica quatro gerações de portais corporativos e considera que, no ano de 1999, os portais disponíveis no mercado passaram da primeira à terceira geração. Além disso, Eckerson acredita que os portais corporativos têm um potencial de expansão ainda maior que os portais públicos [SPO00].

Geração	Categoria	Características
Primeira	Referencial	Máquina de busca, com catálogo hierárquico de conteúdo da web. Cada entrada do catálogo contém uma descrição do conteúdo e um link. Essa geração enfatiza mais a gerência de conteúdo, disseminação em massa das informações corporativas e o suporte à decisão.
Segunda	Personalizado	O usuário, por meio de um identificador e uma senha, pode criar uma visão personalizada do conteúdo do portal, conhecida como "MinhaPágina". Essa visão mostra apenas as categorias que interessam a cada usuário.  O portal pode avisar ao usuário sempre que um novo conteúdo for adicionado às categorias por ele assinaladas. Os usuários podem publicar documentos no repositório corporativo para que esses sejam também visualizados por outros usuários.  Essa geração privilegia a distribuição personalizada de conteúdo.

Terceira	Interativo	O portal incorpora aplicativos, tais como correio eletrônico, calendários, agendas, fluxos de atividades, gerência de projeto, relatórios de despesas, viagens, indicadores de produtividade etc. Essa geração adiciona o caráter cooperativo ao portal, provendo múltiplos tipos de serviços interativos.
Quarta	Especializado	Portais baseados em funções profissionais, para gerência de atividades específicas na instituição, tais como vendas, finanças, recursos humanos etc. Essa geração envolve a integração de aplicativos corporativos com o portal, de forma que os usuários possam executar transações, ler, gravar e atualizar os dados corporativos, e ainda incorpora outras possibilidades como comércio eletrônico, por exemplo.

*Tabela 3.2 – Gerações dos Portais Corporativos*

### 3.3. Tipos de Portais Corporativos

As funções mais importantes de um portal corporativo são suporte à decisão e processamento cooperativo. Alguns autores e fornecedores dão mais ênfase a uma ou outra função e costumam denominar os portais de acordo com sua principal utilização na empresa. O uso de certas nomenclaturas por parte dos fabricantes, entretanto, pode levar a interpretações equivocadas sobre as reais características de seus produtos. Baseado na avaliação da principal utilização do Portal sugere-se uma forma de tipificação que estruture tal conjunto de interpretações.

#### 3.3.1. Portal de Informações

Murray conceitua “Portal de Informações” como aquele capaz apenas de organizar grandes acervos de conteúdo a partir dos temas ou assuntos neles contidos, conectando as pessoas às informações [MUR99]. Nessa categoria Murray inclui as máquinas de busca (internas e/ou externas) e os portais públicos. Nesse tipo de portal não há preocupação com a interatividade e o processamento cooperativo entre usuários e especialistas.

Para Murray, no contexto organizacional, a aplicação de um portal apenas de conteúdo seria insuficiente para atender aos objetivos de negócios das empresas [MUR99]. White

denomina essa forma básica de portal corporativo como Portal Intranet, o qual inclui links para informações e sites web dentro e fora da organização. Assim como Murray, White considera o Portal Intranet análogo aos portais públicos como o Yahoo, por exemplo [WHI01].

### 3.3.2. Portal de Apoio à Decisão

Os portais com ênfase em apoio à decisão auxiliam executivos, gerentes e analistas de negócios a acessar as informações corporativas para a tomada de decisões de negócio. Estes Portais são correspondentes aos sistemas de apoio à decisão – *Decision Support Systems - DSS*. Por darem pouca ou quase nenhuma ênfase ao processamento cooperativo, podem ser incluídos, nessa categoria, o portal de informações de Murray, o portal de negócios de Eckerson e o EIP para processamento de decisões de White. Levando em consideração as características destacadas pela Viador, em seu artigo [VIA00], seu portal de informações empresariais pode ser considerado um exemplo de portal com ênfase em suporte à decisão.

#### 3.3.2.1. Portal de Apoio à Decisão baseado em ferramenta de *business-intelligence*

O portal de apoio à decisão, na concepção de White, permite que “os usuários organizem e encontrem informações corporativas em um conjunto de sistemas que constituem a cadeia produtiva de informações de negócios” [DAI99]. Considerando que o Portal seja estruturado a partir de ferramentas inteligentes (*business-intelligence*) e aplicativos analíticos para capturar informações<sup>2</sup> armazenadas em bases de dados operacionais, no *data warehouse* corporativo ou ainda em sistemas externos à organização e, a partir dessas informações, gerar relatórios e análises de negócio para serem distribuídos eletronicamente aos diversos níveis de tomada de decisão na empresa.

As informações fornecidas por um portal de suporte à decisão podem ser apresentadas sob a forma de relatórios, gráficos e indicadores de desempenho sobre as operações de negócios, podendo ser resumidas ou detalhadas de acordo com o nível estratégico, tático ou operacional do tomador de decisão.

---

<sup>2</sup> As informações manipuladas por esse tipo de portal normalmente são estruturadas e provêm de bancos de dados.

### 3.3.3. Portal de Negócios

Eckerson utiliza o termo portal de negócios como o equivalente corporativo dos portais comerciais da Internet, tais como Yahoo e Excite [PSG00]. Esse tipo de portal tem como função tornar disponíveis, aos usuários corporativos, informações necessárias para a tomada de decisões de negócios da instituição, tais como relatórios, pesquisas, documentos textuais, planilhas, mensagens de correio eletrônico, páginas web, vídeos, dentre outros.

Segundo Davydov, o portal de negócios é o ponto de partida central para os aplicativos de gerência de conteúdo e de processamento de decisões, conectando-os a informações estruturadas e não estruturadas [DAV01]. O Portal deste tipo tem sua ênfase na realização de soluções eletrônicas para o comércio – *e-commerce*.

### 3.3.4. Portal Cooperativo

Os portais com ênfase em processamento cooperativo, assim como os sistemas *groupware* e de automação de escritórios, lidam com informações tanto da cadeia produtiva tradicional, armazenadas e manipuladas por aplicativos corporativos, como informações geradas por grupos ou indivíduos fora dessa cadeia.

Essa categoria de portal, até o momento, foi a que menos se desenvolveu no mercado. Integram essa classe os portais cooperativos de Reynolds & Koulopoulos, os EIPs para processamento cooperativo de White, os portais cooperativos e de especialistas de Murray.

#### 3.3.4.1. Portal Cooperativo baseado em ferramenta de cooperação

Uma especialização do tipo de Portal Cooperativo é o baseado em ferramenta de cooperação. Esse tipo de portal é estruturado a partir de ferramentas cooperativas de trabalhos em grupo (*groupware*) e de fluxo de tarefas/documentos (*workflow*) para prover acesso a informações geradas por indivíduos ou grupos [KST01]. As informações manipuladas por esse tipo de portal são geralmente não estruturadas, personalizadas e encontram-se sob a forma de textos, memorandos, gráficos, mensagens de correio eletrônico, boletins informativos, páginas web e arquivos multimídia. Pode-se caracterizar como uma especialização dos Portais Cooperativos pois este tipo de portal é fortemente apoiado por ferramentas de *groupware* tais como o Lotus Notes. Um produto que apresenta esta característica é K-Station detalhado no tópico 4.4.3-K-Station

### 3.3.5. Portal de Integração de Especialistas

Na concepção de Murray, um portal, para ser completo, deve ter a capacidade de relacionar e unir pessoas com base em suas habilidades e experiências [MUR99]. Essa é a proposta do portal de especialistas - um meio de comunicação e troca de experiências entre pessoas especializadas em determinadas áreas do conhecimento, por meio de comunicação em tempo real, educação à distância e manutenção de cadastro automático de especialistas. Os usuários são agrupados em grupos com um tema específico em comum, também denominados comunidades. A capacidade de integração dos especialistas constitui no elemento principal deste tipo de Portal.

### 3.3.6. Portal de Informações Empresariais

Os portais mais abrangentes, que conseguem aliar as funções de apoio à decisão e processamento cooperativo, conectam os usuários não só a todas as informações, mas também a todas as pessoas necessárias para a realização dos negócios da empresa. Em um mesmo ambiente, são consolidados aplicativos de gestão de conteúdo, processamento de decisões, groupware, workflow, correio eletrônico, inteligência de negócios e sistemas especialistas.

Nessa categoria encontra-se o EIP originalmente definido por Shilakes & Tylman, o EIP de White e o portal do conhecimento de Murray. Considerando as características descritas pelos fabricantes e a possibilidade de conexão a outros aplicativos especializados, pode-se considerar o portal corporativo da Plumtree Software e o EIP da Brio como exemplos dessa categoria. O portal de informações empresariais utiliza metadados e a linguagem XML (*Extensible Markup Language*) para integrar os dados não estruturados, mantidos em arquivos textuais, relatórios, mensagens de correio eletrônico, gráficos e imagens aos dados estruturados das bases de dados do *data warehouse*, fornecendo acesso às informações institucionais a partir de uma interface individualizada, disponível na rede hipertextual corporativa - Intranet.

O EIP alia as características do portal cooperativo e do portal de suporte à decisão. Firestone [EIP00], levando em conta a intenção quase explícita de Shilakes & Tylman de relacionar o EIP a aplicativos cooperativos e interativos para gerenciar repositórios de conhecimento, considera o portal de informações empresariais semelhante ao portal do

conhecimento de Murray. Essa semelhança, porém, depende do grau de cooperação das ferramentas implementadas no EIP e da capacidade de interação entre especialistas, trocando conhecimentos, experiências e habilidades.

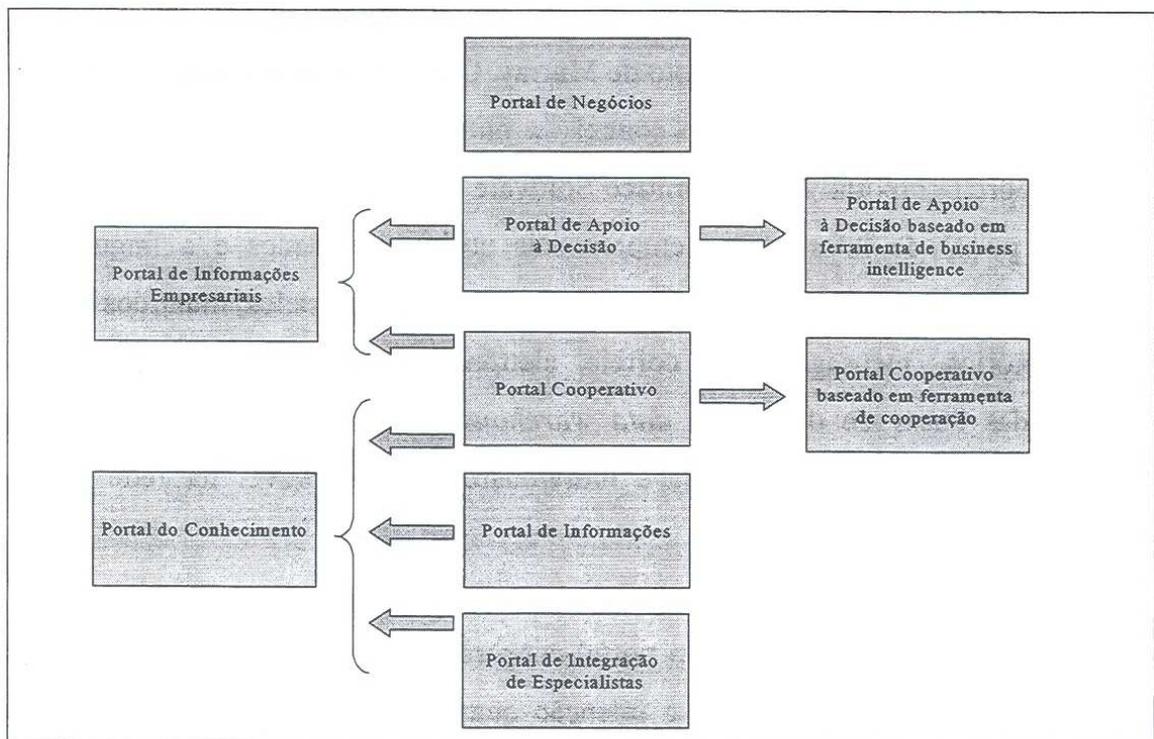
### 3.3.7. Portal do Conhecimento

O portal do conhecimento de Murray, na verdade, é um ponto de convergência dos portais de informações, cooperativos e de especialistas, sendo capaz de implementar tudo que os outros tipos de portais implementam e de fornecer conteúdo personalizado de acordo com a atividade de cada usuário [MUR99].

### 3.3.8. Visão Geral dos Tipos de Portais Corporativos

Considerando os diversos tipos de Portais Corporativos definidos pode-se uma visão resumida (quadro a seguir). Neste quadro resumo fica evidente o núcleo básico de tipos de Portais caracterizado pelas denominações: Portal de Negócios, Portal de Apoio à Decisão, Portal Cooperativo, Portal de Informações e Portal de Integração de Especialistas.

Também podem-se perceber os tipos de portais caracterizados a partir de uma especialização como o Portal Cooperativo baseado em ferramenta de cooperação e o Portal de Apoio à Decisão baseado em ferramenta de business-intelligence.



**Figura 3.1 – Tipos de Portais baseado na utilização**

Por fim, podem-se visualizar os tipos de portais derivados de outros tipos como o Portal de Informações Empresariais que é derivado de Portais de Apoio à Decisão e de Portais Cooperativos e o Portal do Conhecimento que pode ser consolidado a partir dos Portais de Informação, Portais Cooperativos e Portais de Integração de Especialistas. Para evitar confusões de nomenclaturas, será consagrado o conceito Portal Corporativo, em oposição a portal público, já que não há interesse em ressaltar diferenças em relação às funções de cada tipo de portal.

### 3.4. Requisitos mínimos de um Portal Corporativo

Por congregarem, em um só produto, várias tecnologias já existentes, tais como sistemas de inteligência de negócios, gestão de documentos, automação de escritórios, *groupware*, *data warehouse* e *Intranet*, somente para citar os mais relevantes, os fornecedores de produtos nessas áreas têm se posicionado também como fornecedores do mercado de portais corporativos.

Paralelamente, observa-se o aparecimento de pequenas empresas, novas no ramo de informática, que vislumbraram um grande potencial no mercado de portais corporativos, e ainda associações de produtos e empresas em soluções conjuntas para atender necessidades específicas de seus clientes. Com isso, a escolha de um portal corporativo, dentre os vários produtos hoje oferecidos pela indústria de informática, não é uma tarefa tão fácil como parece.

Para auxiliar os executivos das empresas nessa escolha, consultores como Eckerson, do *Patricia Seybold Group*, e White, do *DataBase Associates International*, têm publicado artigos e relatórios contendo requisitos mínimos de um portal corporativo [ECK01, VIA00]. Alguns fornecedores, tais como Plumtree Software, Brio e Viador, têm utilizado esses mesmos relatórios para mostrar a conformidade de seus produtos às regras básicas neles identificadas [BRI00, PSG00].

As 15 regras de Eckerson [ECK01], apresentadas a seguir, resumem os principais requisitos esperados de um portal corporativo.

N.	Requisito	Descrição
I	Fácil para usuários eventuais	Os usuários devem conseguir localizar e acessar facilmente a informação correta, com o mínimo de treinamento, não importando o local de armazenamento dessa informação. Encontrar informações deve ser tão simples quanto usar um navegador web.
II	Classificação e pesquisa intuitiva	O portal deve ser capaz de indexar e organizar as informações da empresa. Sua máquina de busca deve refinar e filtrar as informações, suportar palavras-chave e operadores booleanos, e apresentar o resultado em categorias de fácil compreensão.
III	Compartilhamento cooperativo	O portal deve permitir aos usuários publicar, compartilhar e receber informações de outros usuários. O portal deve prover um meio de interação entre pessoas e grupos na organização. Na publicação, o usuário deve poder especificar quais usuários e grupos terão acesso a seus documentos/objetos.
IV	Conectividade universal aos recursos informacionais	O portal deve prover amplo acesso a todo e qualquer recurso informacional, suportando conexão com sistemas heterogêneos, tais como correio eletrônico, bancos de dados, sistemas de gestão de documentos e <i>groupware</i> . Deve ser capaz de gerenciar vários formatos de dados estruturados e não estruturados.
V	Acesso dinâmico aos recursos informacionais	Por meio de sistemas inteligentes, o portal deve permitir o acesso dinâmico às informações nele armazenadas, fazendo com que os usuários sempre recebam informações atualizadas.
VI	Roteamento inteligente	O portal deve ser capaz de direcionar automaticamente relatórios e documentos a usuários selecionados.
VII	Ferramenta de inteligência de negócios integrada	Para atender às necessidades de informação dos usuários, o portal deve integrar os aspectos de pesquisa, relatório e análise dos sistemas de inteligência de negócios.
VIII	Arquitetura baseada em cliente - servidor	Para suportar um grande número de usuários e grandes volumes de informações, serviços e sessões concorrentes, o portal deve basear-se em uma arquitetura cliente-servidor.

IX	Serviços distribuídos	Para possibilitar o balanceamento da carga de processamento, o portal deve distribuir os serviços por vários servidores.
X	Definição flexível das permissões de acesso	O administrador do portal deve ser capaz de definir permissões de acesso para usuários e grupos da empresa, por meio dos perfis de usuário.
XI	Interfaces externas	O portal deve ser capaz de se comunicar com outros aplicativos e sistemas.
XII	Interfaces programáveis	Deve ser capaz de ser acionado por outros aplicativos, tornando pública sua interface ( <i>API - Application Programming Interface</i> ).
XIII	Segurança	O portal deve prover serviços de segurança como: criptografia <sup>3</sup> , autenticação <sup>4</sup> e <i>firewall</i> . Deve também possibilitar auditoria dos acessos a informações e das alterações de configuração.
XIV	Fácil administração	O portal deve gerenciar todas as informações corporativas e monitorar o funcionamento de forma centralizada e dinâmica. Deve ser de fácil instalação, configuração e manutenção, e aproveitar, na medida do possível, a base instalada de hardware e software adquirida/contratada anteriormente pela organização.
XV	Customização e personalização	Deve possibilitar a customização de acordo com as políticas da organização. Os usuários devem poder personalizar sua interface para agilizar o acesso às informações consideradas relevantes.

**Tabela 3.3 – Requisitos mínimos de um portal corporativo**

A esses requisitos podem ser adicionados ainda algumas proposições de Mark Davydov [DAV01]:

<sup>3</sup> Métodos que permitem esconder o conteúdo de mensagens ou dados armazenados. O processo de cifragem corresponde à transformação da mensagem original em algo ininteligível, utilizando um código secreto - a chave criptográfica [COU94].

<sup>4</sup> Mecanismo de identificação única de um indivíduo. Em sistemas informatizados, normalmente é feita por meio de um identificador e uma senha individual, conhecida apenas pelo usuário [COU94].

- habilidade de gerenciar o ciclo de vida das informações, estabelecendo níveis hierárquicos de armazenamento e descartando as informações ou documentos quando não mais necessários;
- habilidade de localizar especialistas na organização, de acordo com o grau de conhecimento exigido para o desempenho de alguma tarefa;
- habilidade de satisfazer as necessidades de informação de todos os tipos de usuários da organização;
- possibilidade de troca de informações com clientes, fornecedores, revendedores dentre outros, fornecendo uma infraestrutura informacional adequada também para o comércio eletrônico.

### **3.5. Conclusão**

Neste capítulo percebe-se que, a partir dos inúmeros avanços tecnológicos obtidos no mundo *web*, muitas instituições estão sofrendo com o excesso de informações, sendo imprescindível a aplicação da gestão do conhecimento para administrar esse conjunto de dados. A partir da diversidade de alternativas de conceitos apresentados nesta área, busca-se sugerir uma classificação dos Portais, sugerindo-se os principais tipos encontrados. Desta forma, há um direcionamento das principais características que um Portal Corporativo pode apresentar, denotando as especializações e derivações a partir de alguns tipos básicos.

Por fim, apresentam-se os requisitos mínimos de um Portal Corporativo. Este conjunto de características e requisitos define o conceito de Portal Corporativo, demonstrando sua larga abrangência e reforçando a necessidade de suporte a navegação do usuário sobre o qual é focada esta pesquisa. No próximo capítulo será apresentada uma análise das possíveis soluções que podem ser endereçadas para apoiar o processo de busca de conhecimento em Portais Corporativos.

## Capítulo 4

# ANÁLISE DE SOLUÇÕES

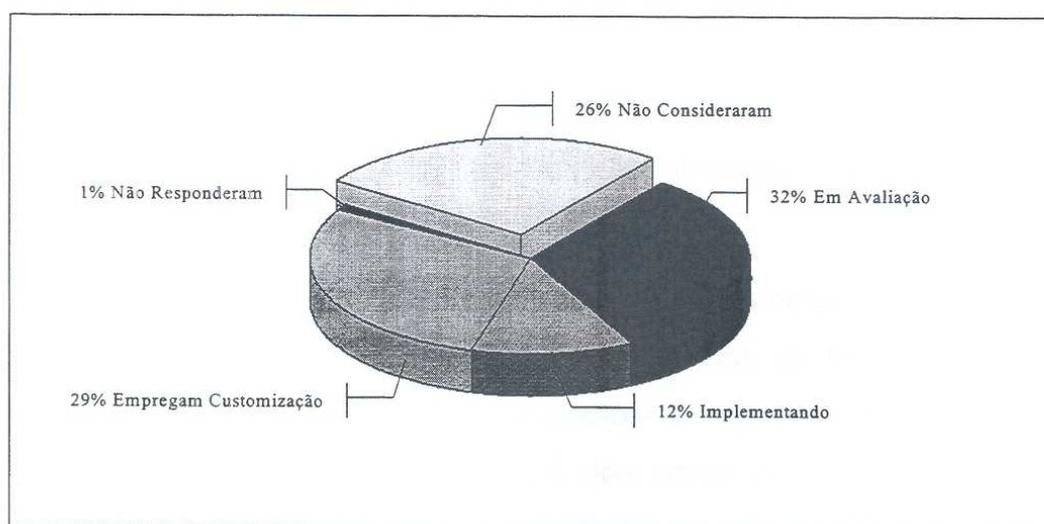
Neste capítulo serão apresentadas as diferentes abordagens para o suporte à navegação dos usuários num site de uma forma geral e, notadamente, num Portal Corporativo. A navegação dos usuários pode ser suportada por algumas ferramentas que indiquem os conteúdos disponíveis. A ferramenta mais difundida é o mecanismo de busca (*search engine*). Atualmente há uma forte tendência pela disponibilização das informações a partir do perfil do usuário, customizando a apresentação das páginas.

### 4.1. Customização de sites

De acordo com um relatório da Forrester Research [INW01], na medida em que o conteúdo genérico dos Web sites torna-se muito grande, os indivíduos irão cessar as atividades de navegar sem rumo e irão gravitar em torno dos sites que ofereçam produtos e serviços customizados para suas necessidades.

Os usuários preferem receber informações que se adaptam as suas necessidades específicas e que sejam correlacionadas as suas atividades diárias. Uma forma de buscar um relacionamento leal e de longa duração com cada indivíduo é com o emprego da customização do site também denominado de personalização. Conforme um estudo do *Jupiter Communications Inc.*, de Setembro de 2000, 34% do total de vendas via web foram dirigidos pela customização e 40% dos portais baseados em negócios eletrônicos (*e-business*) usaram softwares de customização [ZUL01].

De acordo com Fernando Zuliani, consultor da IBM, “a personalização é intrínseca a capacidade de selecionar o conteúdo que melhor se adequa às necessidades e interesses de cada visitante no site”. Zuliani mostra que “numa pesquisa entre fornecedores de Portais mais de 73% estão usando ou avaliando o uso da personalização”. O gráfico a seguir detalha esta pesquisa realizada pela IBM – Fonte: IBM *Market Survey* [ZUL01].



**Figura 4.1 – Distribuição de Fornecedores avaliando Customização[ZUL01]**

O ponto chave é descobrir quais as informações que o usuário realmente necessita no site. Descobrir o que interessa e o que não interessa é um dos princípios básicos da personalização. Como analisar as preferências do usuário? Uma das abordagens empregadas parte do conhecimento obtido a partir dos registros de log do servidor Web.

As ferramentas de customização de Sites conforme as preferências, do usuário podem ser classificadas em duas categorias: **Ferramentas Baseadas em Regras** e **Ferramentas baseadas em Filtragem Colaborativa**.

As ferramentas baseadas em regras são indicadas quando já se conhecem as preferências do usuário, seja a partir de um cadastro prévio ou de informações obtidas em bancos de dados que já tenha as informações cadastradas. Basicamente, tal abordagem utiliza lógicas de programação para indicar qual conteúdo incluir numa página particular. As ferramentas baseadas em filtragem colaborativa são indicadas quando se deseja inferir sobre

as preferências do usuário. Neste caso, são utilizadas técnicas estatísticas complexas para estimar o que é mais apropriado.

## 4.2. Data Webhousing

Como exemplo de ferramentas baseadas em regras pode-se caracterizar o modelo concebido a partir da metodologia de *Data Webhousing* proposta por Kimball [KIM00] através da análise dos logs do servidor de Web. Em uma primeira fase são apurados os principais itens de informação obtidos dos logs com vista a compor um *Data Mart* e posteriormente avaliam-se novas iniciativas com vista a melhorar tanto a coleta de dados como melhorias gerenciais e operacionais. Unir a Web ao *Data Warehouse*, significa atribuir comportamento ao *Data Warehouse* [KIM00]. É o caso de customização a partir de regras

O consórcio entre as tecnologias *Web* e *Data Warehouse* originou o que se conhece por *Data WebHousing*. Segundo [KIM00], *Data WebHousing* é uma evolução natural do *Data Warehouse* e possui três grandes desafios de projeto:

- **Temporalidade** - Os resultados do negócio precisam estar disponíveis em tempo real.
- **Volume de Dados** - enorme capacidade de armazenamento de dados<sup>5</sup>. Cada ação, representada por um clique do mouse, efetuada por um consumidor deverá ser armazenada para que possa ser analisada à posteriori.
- **Tempo de Resposta** - O tempo de resposta à uma solicitação do usuário deverá ser o menor possível.

Um conceito fundamental que deve ser ressaltado é o "*clickstream*". O *clickstream* é representado pela série de ações que podem ser capturadas durante uma sessão iniciada por um usuário. É através desta trajetória de ações, que são inicialmente armazenadas nos logs de um servidor de Web, que é possível analisar o comportamento e interesses de cada usuário que acessa o Web site da corporação. O *clickstream* é a conexão entre o usuário da Web e o ambiente de negócio da corporação. Uma outra definição proposta por William H. Inmon

---

<sup>5</sup> Em Julho/1998 o número de servidores Internet foi estimado em 125,9 milhões – Internet Domain Survey [IDS02]

[INM00]: Os registros *clickstream* são artefatos detalhados deixados por um usuário da Internet através de suas atividades ou transações que ocorreram durante um diálogo.

Resumindo, *clickstream* é uma imensa fonte de dados atemporal e não disciplinada que registra todas as interações de um visitante qualquer com um Web site. Esta fonte de dados tem um potencial de informação muito maior que qualquer outra fonte tradicional, pois captura as ações e decisões de um visitante que serão carreadas do ambiente Web para o ambiente de Data Warehouse.

A quantidade de dados presente no log do servidor de web é muito grande, o que poderia criar um *Data Mart* de *clickstream* gigantesco. Faz-se necessário delimitar os dados a serem analisados, sendo fundamental traçar objetivos à priori, a fim de viabilizar o trabalho de análise do *clickstream*. Vale ressaltar que os dados, quando desacompanhados de suas definições, ou seja, de seus metadados, não possuem significado. A partir dos dados é necessário obter informação através de um processo coerente de organização e de identificação de padrões. A partir deste momento podem ser feitas correlações, comparações e análises de causa e efeito.

De acordo com [KIM00] os tipos de decisão possíveis quando se identifica e reconhece os clientes são:

- Personalização das atividades de marketing.
- Foco das atividades de marketing através da escolha do grupo de clientes desejado.
- Agregação de clientes recém-chegados a grupos de clientes com perfis comportamentais definidos.
- Detecção em tempo real da intenção do cliente em finalizar a sua visita ao Web site e indicação de alternativas.

De acordo com [KIM00], alguns tipos de decisão são possíveis em relação a comunicação com os usuários, conforme segue:

- Detecção quando anúncios do Web site surtem o efeito desejado, ou seja, taxa de eficiência (efeitos causais).

- Personalização de acordo com características dos usuários;
- Eficiência do Web site, como adequação da interface, disposição da informação, gráficos e opções, escolha de assuntos, quantidade de telas em cascata para transações e conteúdo indicado de maneira apropriada.
- Estímulo ao senso de comunidade, com fóruns de discussão ou informações orientadas a um certo público específico;

Com vistas a aumentar a interação com os clientes, os conceitos enumerados acima deveriam ser avaliados de forma a determinar quais seriam os mais adequados à proposta do Web site.

### **4.3. Web Mining**

A mineração de dados é uma das etapas do processo de descoberta de conhecimento em bancos de dados e têm por objetivo extrair informações implícitas e potencialmente úteis de dados [FAY96]. Normalmente todo este processo consiste em uma série de etapas, que inicia com a seleção do conjunto ou amostra dos dados a serem processados caracterizando um processo baseado em filtragem colaborativa.

Estes dados podem ser submetidos a uma etapa de pré-processamento, onde são transformados para um formato adequado para o algoritmo de mineração, que procura por regras ou padrões ocultos nos dados. Finalmente as informações descobertas são selecionadas e formatadas para a exibição, afim de serem interpretadas e avaliadas, de forma que se selecione os conhecimentos úteis resultantes de todo o processo. Dentre as diversas técnicas de mineração de dados, uma que se destaca é a procura por regras de associação, que têm por objetivo encontrar relacionamentos ou padrões freqüentes entre conjuntos de dados [AGR96].

A abordagem de Web Mining parte da hipótese de que estas técnicas podem ser utilizadas para auxiliar a compreensão do comportamento do usuário da Web, através da descoberta de padrões ou regularidades que descrevam o seu perfil. Estas informações poderiam auxiliar no projeto ou remodelagem dos sites, no gerenciamento dos recursos dos provedores de conteúdo na Web, no desenvolvimento de agentes que auxiliem a navegação ou o aprendizado do usuário, entre outras finalidades.

Há disponível uma grande quantidade de ferramentas que, através da análise do arquivo de *log* dos servidores Web, fornecem informações estatísticas sobre os acessos às páginas [ALA00], como por exemplo, número de acessos por páginas, tráfego total e origem das requisições.

O Web Mining, que pode ser resumido como o uso de técnicas de data mining para descobrir e extrair de forma automática, informação dos documentos e serviços existentes na Web, pode resolver algumas tarefas, conforme Etzioni [ETZ96]:

- Procura de recursos (*Information Retrieval*) : obter os documentos desejados da Web.
- Seleção e pré-processamento de informação (*Information Extraction*): seleção e pré-processamento automático de informação obtida da Web.
- Generalização: descoberta automática de padrões em Web sites individuais ou em sites múltiplos.
- Análise: validação e/ou interpretação dos padrões descobertos.

Conforme Cooley [COO97], o uso de mineração de dados no ambiente da Web pode ser sub-dividida em duas categorias: a Mineração de Conteúdo da Web (*Web Content Mining*) e a Mineração de Uso da Web (*Web Usage Mining*).

#### **4.3.1. Mineração de Conteúdo da Web**

Define-se Mineração de Conteúdo da Web (*Web Content Mining*) como sendo a descoberta de informação útil a partir do conteúdo e/ou dados e/ou documentos da Web. Seu escopo é centrado na diversidade destes elementos (conteúdos, dados e documentos) e na análise de sua ocorrência na Web. Também pode ser identificado como mineração de textos na Web (Text mining).

Este primeiro caso é implementado com a busca automática dos recursos e recuperação de informações disponíveis na rede. Como exemplo desta abordagem, pode-se citar as ferramentas de busca, como Altavista e Google, entre outros.

### 4.3.2. Mineração de Uso da Web

O processo de Mineração de Uso da Web (*Web Usage Mining*) tem como prerrogativa tentar prever e/ou explicar o comportamento dos usuários que interagem com a Web. Uma das principais fontes de dados para este tipo de extração de dados são os logs criados pelos servidores de Web. A restrição a esta abordagem é o fato de que a informação registrada nos logs pode não refletir com perfeição a navegação do site, devido aos vários níveis de caching que existem na Web. Outros tipos de informação, como por exemplo a informação enviada pelo método POST dos formulários, não é registrada nas logs.

A mineração de uso da Web consiste no uso de ferramentas que automaticamente descubram padrões de acesso dos usuários. Zaiane [ZAI98] aplicou técnicas de OLAP (*Online Analytical Processing*) sobre o log, onde através de classificações e sumarizações foram encontradas algumas correlações entre eventos e analisadas séries temporais. Nasraoui [NAS00] descreve um algoritmo para a aglomeração dos acessos em seções típicas, enquanto que Spiliopoulou [SPI00] construiu árvores agregadas que representam seqüências freqüentes de acesso, sobre a qual foram aplicadas técnicas de mineração.

### 4.4. Ferramentas de Mercado

Os portais aceleram a transformação de informação em inteligência para o negócio, porque eles agem como lentes entre as diversas fontes de dados e as ferramentas de gestão de conhecimento, que filtram detalhes específicos. Depois de selecionados dados com alguma relação, os portais refinam as informações mais relevantes, selecionando dados significantes no cenário da empresa, em seguida, a gestão do conhecimento refina ainda mais estes dados, detectando padrões relevantes nas informações que levam à inteligência do negócio.

Os portais fornecem uma solução para as práticas de gestão do conhecimento em um único *front-end*. Para as organizações, o desafio está em determinar como juntar todos os pedaços tecnológicos necessários, fonte de dados e serviços de gestão do conhecimento em uma solução que forneça valores que possam ser realmente medidos. Algumas ferramentas simplificam o processo de construir o portal e incorporar os vários componentes da gestão do conhecimento necessários ao negócio. Neste tópico serão analisadas algumas destas

ferramentas, particularmente no que tange a capacidade de auxiliar o usuário no processo de navegação do Site.

#### **4.4.1. MyLivelink Portal**

A Open Text Corp. é uma empresa de Gestão do Conhecimento que oferece uma ferramenta para implantação de portal. O seu produto, MyLivelink [MLL01], permite que os usuários customizem seus browsers para acessar as funcionalidades do Livelink (produto baseado na web para colaboração), assim como os dados armazenados nos repositórios do Livelink. As funcionalidades incluem lista de tarefas, discussões (*chat*) e novos canais (novidades, esportes, informações da bolsa). Com as listas de tarefas, gerentes e trabalhadores passam tarefas entre eles, para que elas sejam assinadas, acompanhadas e aprovadas.

MyLivelink foi desenvolvido também para facilitar o gerenciamento de conteúdo – usuários podem coletar conteúdo em uma página constituída de várias áreas de um sistema Livelink ou a partir de vários sistemas Livelink. Para funcionalidades internas, são usadas mini APIs que permitem que sejam integradas rapidamente funções do Livelink como busca e recuperação. Para aplicações externas, são utilizadas APIs similares, que integram aplicações externas como SAP e Lotus Notes mais facilmente dentro do myLivelink.

MyLivelink oferece 3 visões do ambiente de trabalho: Empresa, projeto e pessoal. A visão empresa permite que os empregados colaborem nos dados e aplicações baseados na intranet. A visão projeto permite que os membros dos projetos colaborem virtualmente como uma equipe. Na visão pessoal, cada trabalhador utiliza o portal para suas responsabilidades específicas. MyLivelink integra o gerenciamento de documentos e workflow. O produto é flexível, fácil de administrar e usar e altamente escalonável.

#### **4.4.2. Plumtree Corporate Portal 4.0**

A Plumtree [PLU01] introduziu a prática de disponibilizar as funcionalidades de aplicações externas através de módulos. Para aumentar a velocidade de disponibilização de portais de clientes, a Plumtree oferece mais de 600 módulos certificados para aplicações como o CRM da *Siebel* e o ERP da SAP.

Depois de usar o índice para mapear as classificações de páginas, o *Plumtree Corporate Portal* carrega automaticamente, filtrando o conteúdo em diferentes categorias para serem editadas pelos gerenciadores de conteúdo. De acordo com o seu perfil no portal, os usuários são notificados por email, voz ou celular das atualizações dos conteúdos relevantes a eles. Se os usuários não usarem o conteúdo de algumas fontes, o portal para de enviar as notificações para estes usuários.

O Portal Corporativo Plumtree pode suportar até 60.000 usuários e 3,7 milhões de *hits* por servidor. Todos os objetos do portal podem ser controlados quanto ao acesso a partir de um sistema de autenticação que também garante o acesso ao repositório. Desta forma, os administradores também podem dar direitos de postagem para certos usuários.

#### **4.4.3. K-Station**

K-Station é um portal de conhecimento projetado para possibilitar aos usuários obter um único ponto de acesso para informações críticas, desenvolvido pela Lotus Development Corporation [KST01]. O Lotus K-station, também denominado de Raven, é uma combinação das tecnologias de Portal da Lotus bem como ferramentas e serviços de cunho colaborativo tais como Sametime and QuickPlace.

A Lotus Raven permite a classificação de informações de forma inteligente – após selecionados os repositórios a serem carregados, O processo de classificação do Raven passa por um estágio de rascunho da classificação, onde os documentos com conteúdos similares são agrupados. Tais documentos podem ser customizados posteriormente de acordo com a preferência do usuário. Adicionalmente, a ferramenta tem a capacidade de aprender as preferências definidas pelo usuário.

A ferramenta Raven possui um localizador de Especialista, que agrupa documentos de autores com conteúdos no qual eles são especialistas. Assim os usuários podem explorar o conhecimento destes especialistas de forma organizada e agrupada. Raven possui também um conjunto de módulos de colaboração, em especial o QuickPlace. Este módulo possui a facilidade “*awareness*” que alerta ao usuário quem está online em cada espaço cooperativo, para que se possa efetivar atividades instantâneas como reuniões virtuais.

#### 4.4.4. Broadvision

A ferramenta *BroadVision One-To-One Content* [BRO01] é uma solução de gerenciamento de conteúdos escalável que possibilita às organizações maximizar o valor de seus ativos baseados em informações ou, notadamente, em conteúdos.

O *BroadVision One-To-One Content* é uma aplicação standalone que pode ser integrada em diversos ambientes para o gerenciamento de conteúdos da organização. Conteúdos endereçados como “tags” facilitam a personalização com qualquer mecanismo de distribuição que selecione dinamicamente tais informações a partir de requisições do usuário.

A solução *Broadvision* apresenta uma arquitetura robusta baseada em banco de dados escalável a partir da habilidade de adicionar servidores de conteúdo para atender a usuários adicionais. Uma arquitetura aberta suportando os padrões de mercado XML e Java J2EE possibilita a implementação de funções estendidas a partir de APIs que possibilitam customizar procedimentos de workflow além da interface do usuário.

#### 4.4.5. Websphere Personalization

O *Websphere Personalization* [WSP01] permite construir sites na Web e correlacionar os conteúdos com as necessidades e interesses de cada usuário, incrementando a produtividade dos funcionários no caso de soluções *B2E – Business-to-employee*. As principais funcionalidades apresentadas são:

- **Controle de Personalização** — ferramenta de apoio a usuários finais (não técnicos) que possibilita direcionar os objetivos do negócio.
- **Gerenciamento de Campanhas** — permite ao usuário final definir uma campanha no site enfatizando a apresentação de determinado conteúdo. Função integrada com o controle de personalização.
- **Perfil Implícito** — possibilita descrever o perfil de visitantes do site e personalizá-lo baseado nas ações desenvolvidas pelo usuário.
- **Relatórios de Efetividade** — controles gerenciais de efetividade das campanhas lançadas.

- **Regras de Negócio Adicionais** — possibilita a filtragem de e-mails para um determinado usuário.

#### 4.5. Conclusão

Neste capítulo procurou-se apresentar as principais estratégias de suporte à navegação dos usuários num site de uma forma geral e, notadamente, num Portal Corporativo. Em especial, foram analisadas as possibilidades de customização de um Portal possibilitando auxiliar o processo de busca de conhecimento. Como elemento de registro de conhecimento foi reforçado o log do servidor web que possibilita uma análise das páginas acessadas.

Também foi apresentado que as diversas ferramentas de mercado possuem características poderosas de controle e personalização dos sites produzidos a partir de suas soluções. Há uma clara preocupação pelo gerenciamento dos conteúdos e das informações, possibilitando agregar valor ao processo de navegação do usuário. A tônica das ferramentas é na disseminação de políticas de *marketing one-to-one* onde os processos de personalização empregam regras pré-definidas para estabelecer uma possível sugestão de informação para o usuário.

Via de regra, não há preocupação em suportar a navegação do usuário a partir do aprendizado obtido com outros usuários que já tenham navegado anteriormente por aquele site. A possibilidade de agregar valor ao conhecimento obtido por um usuário não é explorada por estas ferramentas de mercado. Há um potencial de uso deste tipo de conhecimento exatamente a partir da análise da log do servidor web. Em outras palavras, extraindo informações a partir da log é possível identificar o conhecimento de um determinado grupo de usuários que navegam diretamente para os documentos relevantes sem passar por outras páginas não diretamente relacionadas. Assim, é possível estruturar uma solução computacional que sugira o próximo passo, para um usuário sem conhecimento de um determinado conjunto de informações.

No capítulo seguinte é apresentada uma proposição de emprego desta premissa – de que é possível identificar o conhecimento de navegação de um grupo de usuários sugerindo as páginas corretas para um usuário sem conhecimento a cerca deste conjunto de informações.

Procura-se, ainda, focar esta proposição para o emprego em Portais Corporativos com ênfase em portais da indústria financeira, onde há uma prioridade em soluções que agreguem valor e possibilitem redução de custos tanto com aprendizado como no rápido e eficaz acesso aos dados.

## Capítulo 5

# SUORTE A NAVEGAÇÃO BASEADO NO PERFIL DO USUÁRIO

Em um recente relatório apresentado pelo *Gartner Group* afirma-se que em 2003 mais da metade das grandes corporações mundiais vão implantar sistemas de gestão do conhecimento. O instituto aposta ainda que 60% das interações com clientes serão eletrônicas em 2004 e que as empresas serão capazes de prever as preferências dos clientes. E, em 2010, mais de 75% das análises de respostas de consumidores serão automatizadas.

Conforme artigo do Information Week [INW01]: "As máquinas vão responder aos usuários e encaminhar as medidas para as providências serem tomadas", avalia Pedro Bicudo, vice-presidente de consultoria do Gartner no Brasil. De acordo com Bicudo, em 2008 será o reinado do ERP II, com novos pacotes integrados que irão cobrir todas as áreas de parcerias de negócios, incluindo fornecedores e canal de distribuição, além de ferramentas de business intelligence (BI) e gestão do conhecimento. "O ERP II será especializado por setor", explica.

Buscando atender a esta tendência de mercado e consolidando os diversos conceitos apresentados neste trabalho, será focado este estudo na definição de uma ferramenta de suporte à navegação em um Portal Corporativo, que subsidie a gestão de conhecimento, a partir da avaliação do comportamento dos usuários.

## 5.1. Arquitetura

Cada produto disponível no mercado de portais corporativos tem características próprias, estrutura diferenciada ou componentes adicionais, apresentadas como vantagens competitivas, quando comparado aos concorrentes. Entretanto, a arquitetura básica de qualquer portal corporativo segue, em linhas gerais, o modelo descrito por White [DAI01], composto por um assistente de informações, provido por um navegador web (*browser*) e um servidor web com os seguintes elementos: diretório de informações de negócios, máquina de busca, analisador de metadados, ferramenta de publicação, ferramenta de assinatura, interfaces de importação e exportação de dados.

Os portais corporativos disponíveis no mercado podem ser produtos independentes ou produtos integrados em outros softwares, como é o caso de portais embutidos em *groupwares*, em ferramentas de inteligência de negócios ou em softwares de gestão de documentos.

A proposta deste trabalho é estabelecer um banco de informações que consolide o controle de log de servidores Web, permitindo trilhar as características de acesso dos usuários, notadamente no que se refere aos Portais Corporativos. Considerando o endereço IP como chave referencial do usuário, pretende-se estabelecer o padrão de comportamento de navegação deste usuário, possibilitando enfatizar as informações mais acessadas e revisar o conteúdo de páginas com baixa frequência de acesso.

Esta solução estará de acordo com a proposta de portal defendida por Reynolds & Koulopoulos [REY00] quando identificam que as fases mais avançadas do progresso do portal web correspondem a personalização e ao uso de funções expandidas tais qual a sugerida neste trabalho.

Também se pode montar um paralelo com a definição de Portal de Informações Empresariais utilizado por Shilakes & Tylman: “Portais de Informações Empresariais são aplicativos que permitem às empresas libertar informações armazenadas interna e externamente, provendo aos usuários uma única via de acesso à informação personalizada necessária para a tomada de decisões de negócios” [SHI99].

Para caracterizar os grupos de usuários que apresentam o mesmo perfil de navegação, considera-se que os usuários de um mesmo departamento ou com foco nos mesmos serviços, deverão demandar informações do Portal que sejam similares. Assim, o comportamento na navegação do site deverá ser similar para estes profissionais, caracterizando uma mesma trilha de navegação. Uma vez identificada esta trilha padrão, será possível atuar de forma pró-ativa, sugerindo novas páginas correlacionadas a partir de comportamentos de navegação similares apresentados pelos usuários do site.

Conforme Silberschatz [SIL96], as medidas de interesse dos padrões descobertos no processo de descoberta de conhecimento podem ser tanto objetivas como subjetivas. As medidas objetivas levam em conta a estrutura do padrão encontrado além dos dados utilizados, enquanto as medidas subjetivas levam em conta também o usuário que analisa os dados, isto é, o que pode ser interessante para uma pessoa pode não ser para outra. Na ferramenta proposta, busca-se integrar ambas as medidas de interesse propostas por Silberschatz em seu trabalho.

## 5.2. Requisitos a serem contemplados

Considerando os diversos requisitos propostos por Eckerson [ECK01] conforme visto no item 3.4 – Requisitos Mínimos de um Portal Corporativo, podem-se identificar 3 requisitos que poderão ser contemplados pela ferramenta de suporte à navegação na Web. São eles:

N.	Requisito	Descrição
I	Fácil para usuários eventuais	Os usuários devem conseguir localizar e acessar facilmente a informação correta, com o mínimo de treinamento, não importando o local de armazenamento dessa informação. Encontrar informações de negócios deve ser tão simples quanto usar um navegador web.
V	Acesso dinâmico aos recursos informacionais	Por meio de sistemas inteligentes, o portal deve permitir o acesso dinâmico às informações nele armazenadas, fazendo com que os usuários sempre recebam informações atualizadas.

VI	Roteamento inteligente	O portal deve ser capaz de direcionar automaticamente relatórios e documentos a usuários selecionados.
----	------------------------	--

**Tabela 5.1 – Requisitos mínimos da ferramenta de suporte à navegação**

De uma forma geral os demais requisitos não são contemplados por esta ferramenta de suporte à navegação. Tão importante quanto apontar os requisitos contemplados é demonstrar os que não foram contemplados com o respectivo motivo, conforme segue:

N.	Requisito	Motivo
II	Classificação e pesquisa intuitiva	A ferramenta não indexa as páginas a partir de seu conteúdo.
III	Compartilhamento cooperativo	A ferramenta aprende o perfil de navegação a partir dos registros do log e não possibilita que os usuários cadastrem seu comportamento.
IV	Conectividade aos recursos	Não há vínculo com outras aplicações.
VII	Ferramenta de inteligência	Não é definida uma interface com ferramentas de <i>business-intelligence</i> .
VIII	Arquitetura cliente - servidor	Este requisito poder-se-ia classificar como contemplado, no entanto, entende-se que a ferramenta de suporte à navegação é apenas um componente do Portal.
IX	Serviços distribuídos	Idem ao anterior.
X	Permissões de acesso	A ferramenta não controla acesso sendo disponibilizada para qualquer usuário (usuário <i>guest</i> ).
XI	Interfaces externas	Não foram analisadas interfaces com outras ferramentas. Pode ser apontado como referência de continuação da pesquisa
XII	Interfaces programáveis	Idem ao anterior
XIII	Segurança	Idem ao requisito X (não controla acesso)

XIV	Fácil administração	Idem ao requisito VIII (não controla acesso)
XV	Customização e personalização	Não há possibilidade de customizar a ferramenta. Este tópico pode ser apontado como referência de continuação da pesquisa.

**Tabela 5.2 – Requisitos não contemplados**

A ferramenta de suporte à navegação em Portais Corporativos deverá prover um mecanismo de fácil utilização, sugerindo ao usuário as próximas páginas para sua navegação, considerando seu perfil de utilização e seu grupo de atuação. A solução proposta permite às empresas coletar dados de seus usuários na medida em que as páginas vão sendo navegadas. Desta forma a empresa passa a ser capaz de apresentar suas informações e aplicações no Portal Corporativo com muito mais produtividade. A partir desta solução é possível traçar um mapa da navegação padrão dos usuários e identificar como é o perfil de utilização dos serviços disponíveis, avaliar a eficiência das campanhas de marketing, ou seja, definir o padrão de comportamento dos usuários frente ao site.

Cada clique irá gerar um registro na log do servidor web que, uma vez coletado, irá alimentar um banco de dados transacional com as informações que identificam as dimensões básicas: tempo, local, usuário. A coleção destes registros mostra o “rastro” deixado pelo usuário no site e permite uma análise tanto do ponto de vista do usuário como do ponto de vista estatístico. A combinação destas informações com os dados transacionais da empresa gera uma capacidade analítica extremamente poderosa que pode ser utilizada para produzir decisões estratégicas importantes nas empresas.

É comum denominar-se esta combinação de *Webhousing* ou *E-Intelligence*. A previsão feita por especialistas é que este será o próximo passo das estratégias de negócio baseadas na web: Aproximar-se cada vez mais do usuário, particularmente do funcionário, para facilitar suas atividades diárias é fator preponderante nos projetos de Portal Corporativo.

### **5.3. Resultados Esperados pelo Modelo Proposto**

Considerando os tópicos acima apresentados, seguem os principais resultados esperados:

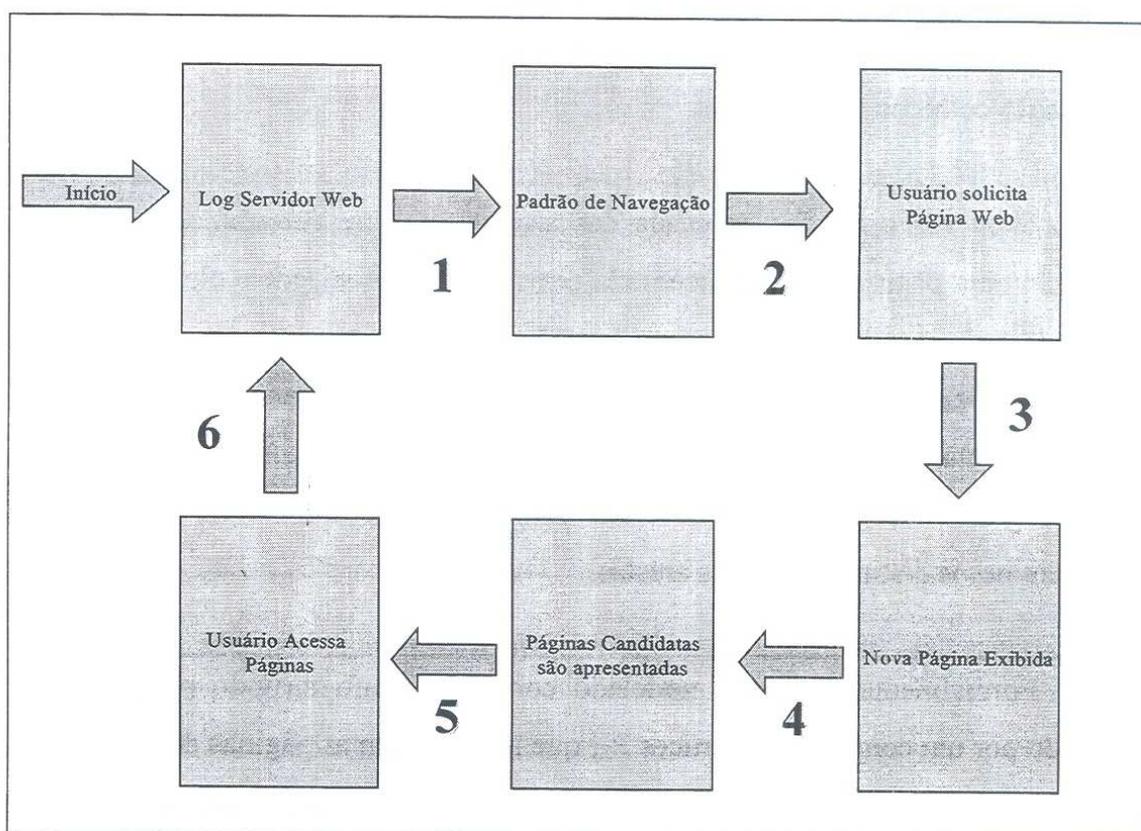
- **Definição de Tecnologia** - o processamento de logs de servidores web requer o conhecimento de uma tecnologia que incorpora características inovadoras de software. No mercado há uma forte busca por diferenciais dos diversos fornecedores visando a conquista deste mercado. Tais pesquisas tem suscitado inúmeras alternativas originais e revolucionárias. O emprego de logs para a solução de acesso a páginas considerando o perfil do usuário, possibilita um grande diferencial de abordagem para as soluções atualmente apresentadas.
- **Apoio a Decisão** - é um conceito corrente de que os sistemas de apoio a decisão (*Decision Support Systems - DSS's*) tem cada vez mais sido apontados como ponto fundamental para as organizações. Também é sabido que, para obter-se informações importantes para os DSS's, é fundamental a manutenção de imensos bancos de dados (centenas de Gigabytes). Conforme visto, o Portal Corporativo pode ser visto analogamente como um data-warehouse onde estão disponíveis milhares de páginas de informações. O emprego da gestão do conhecimento, com a adoção de ferramentas de *business intelligence* – tais como a solução sugerida neste trabalho - permite apresentar soluções eficazes para estes casos.
- **Predição (Predictibilidade)** – característica que permite “predizer” a necessidade de informações por parte dos usuários, considerando seu comportamento de navegação na web em comparação com o comportamento médio de outros usuários com características similares.
- **Personalização** – propriedade que permitirá ao usuário identificar seu browser como seu ambiente de trabalho (*personal desktop*) onde as diversas soluções computacionais estarão dispostas conforme suas necessidades e, adicionalmente, sua navegação estará sendo rastreada para possibilitar consultas de endereços futuros de páginas.

#### 5.4. Visão Geral da Solução

Na figura abaixo apresenta-se uma visão geral da solução. O processo denota-se cíclico, possibilitando uma retroalimentação com o próprio comportamento derivado dos

padrões sugeridos. Num primeiro passo é realizada uma leitura da log do servidor web. Na sequência, avalia-se o padrão de navegação dos usuários. Tal padrão de navegação é registrado para consulta posterior.

A medida que o usuário navega pelo Portal Corporativo, solicitando novas páginas, um *servlet* solicita uma avaliação do padrão de possíveis novas páginas para aquele usuário e para aquele comportamento de navegação.



**Figura 5.1 – Visão geral do processo de suporte a navegação em um Portal Corporativo**

Por fim, o usuário opta a partir das páginas candidatas gerando um novo registro no log do servidor web que irá realimentar o banco de informações de navegação sugerindo um aprimoramento do comportamento de navegação no Portal.

## 5.5. Projeto Físico

Este item detalha as características técnicas de uma ferramenta de suporte à navegação em Portais Corporativos a partir da análise do comportamento de navegação dos usuários.

Esta ferramenta recebe como entrada os registros de log do servidor web e organiza seus dados de forma que seja possível extrair informações como o perfil de determinado grupo de usuários a partir da análise da seqüência de páginas acessadas durante um determinado período.

Inicialmente o site é modelado como um grafo dirigido, no qual os vértices representam as páginas e as arestas representam os links entre as páginas. A cada aresta associa-se um peso relativo inicial igual a zero. A partir desse modelo inicial, a ferramenta de análise de registros de log incrementa os pesos das arestas visitadas, com base nos dados de navegação colhidos nos registros.

Uma vez ajustados os pesos das arestas para um determinado grupo, é possível, a partir de uma página qualquer acessada por um usuário, sugerir a ele a próxima página mais acessada pelo seu grupo. Esta funcionalidade é implementada na forma de um programa executado pelo navegador do usuário, que detecta as mudanças de páginas do mesmo e apresenta então as sugestões adequadas. Conclui-se, portanto, que cada grupo possuirá seu próprio grafo. Os grafos dos diversos grupos compartilham uma mesma topologia, mas possuem pesos distintos em suas arestas.

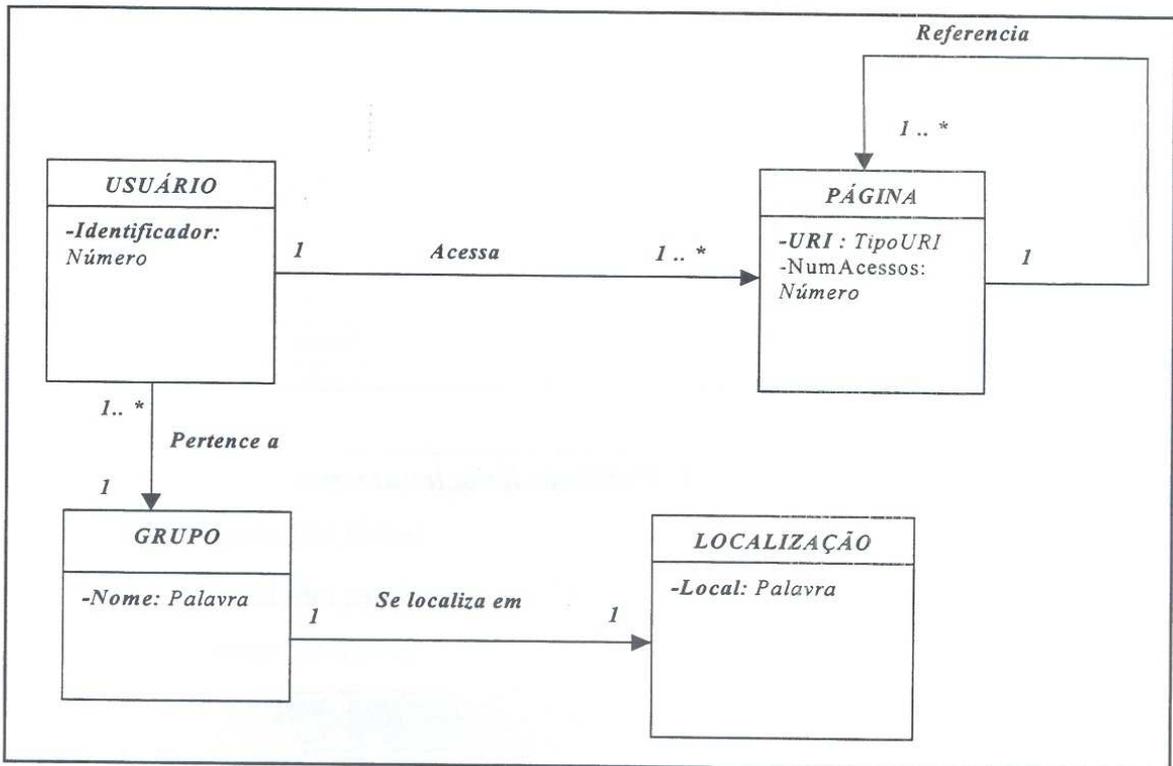
Formalmente, o site é modelado como um grafo dirigido ponderado  $S = (P_S, L_S)$  definido por um conjunto de vértices  $P_S$ , que representam as páginas do site, e um conjunto de arestas  $L_S$ , que representam seus links. Cada link  $l_i \in L_S$  é uma tripla  $[o_i, d_i, p_i]$ , onde  $o_i, d_i \in P_S$  são respectivamente as páginas de origem e de destino do link  $l_i$ , e  $p_i$  seu peso relativo. Links onde  $d_i \notin P_S$  não são considerados. Inicialmente  $p_i = 0 \forall l_i \in L_S$ .

Definindo-se  $ta(p_k)$  como o instante de acesso a uma página  $p_k \in P_S$ , considera-se que um usuário visitou um link  $l_i \in L_S$  se  $ta(d_i) - ta(o_i) \leq \Delta_{max}$ , onde  $\Delta_{max}$  representa o período de inatividade máximo entre duas páginas em uma navegação no site. Intervalos de acesso maiores que  $\Delta_{max}$  serão considerados navegações independentes. O item 5.6.6. Determinação do Período de Inatividade Máximo apresenta um detalhamento desta especificação.

Ao analisar os registros de log do servidor Web, a ferramenta incrementa os pesos dos links visitados (nos quais  $ta(d_i) - ta(o_i) \leq \Delta_{max}$ ). Desta maneira pode-se definir o perfil de um grupo como sendo o conjunto de *passeios* no grafo cujas arestas possuem o maior peso. Cabe lembrar que um passeio em um grafo é uma seqüência de vértices  $p = (v_0, v_1, v_2, \dots, v_n)$ , onde  $v_0, v_1, \dots, v_n \in P_S$  e  $\{v_{i-1}, v_i\} \in L_S$  para todo  $0 < i \leq n$ . O tamanho (comprimento) de  $p$  é denotado por  $|p|$ . Um passeio pode ser uma *trilha*, se todas as suas arestas são distintas, *fechado* se  $v_0 = v_n$ , ou um *circuito*, se  $v_0 = v_n$  e todas as suas arestas são distintas.

## 5.6. Modelo Conceitual

Considera-se que um usuário possui uma identificação única e pertence a determinado grupo. Cada grupo possui uma localização fixa. As páginas são identificadas através da URL, que é composta pelo caminho completo mais o nome do arquivo, conforme apresentado na Figura 5.4 - Modelo Conceitual.



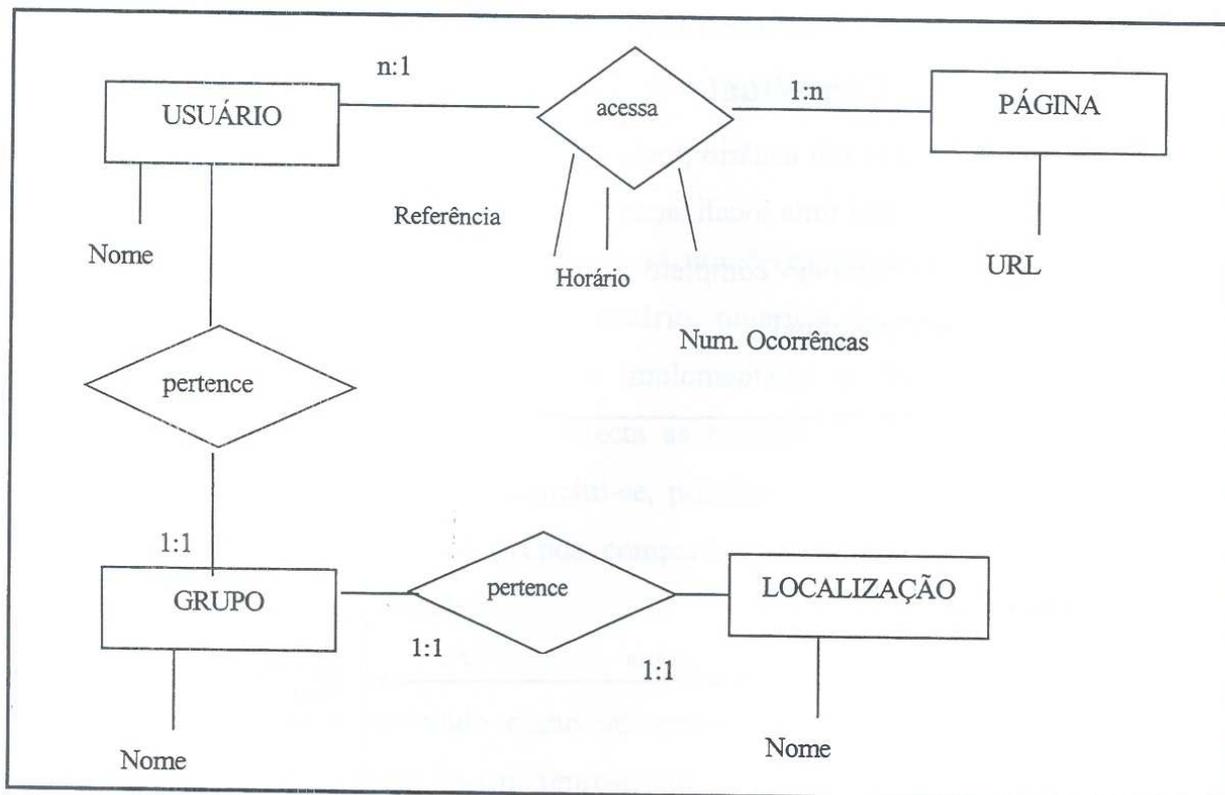
*Figura 5.2 – Modelo Conceitual*

A localização é considerada fixa pois a solução é dirigida para Portais Corporativos onde os endereços IPs dos microcomputadores dos funcionários são dedicados e não há uso

compartilhado de recursos. Assim, um funcionário interage no site a partir de um único endereço (fixo). Adicionalmente, os funcionários estão relacionados a um único departamento, caracterizando o grupo do usuário.

### 5.6.1. Modelo de Dados

Dado o Modelo Conceitual apresentado, deriva-se o seguinte modelo entidade-relacionamento das principais entidades utilizadas nesta ferramenta:



*Figura 5.3 – Modelo de Entidade-Relacionamento*

Tal modelo tem como principais entidades, as seguintes tabelas:

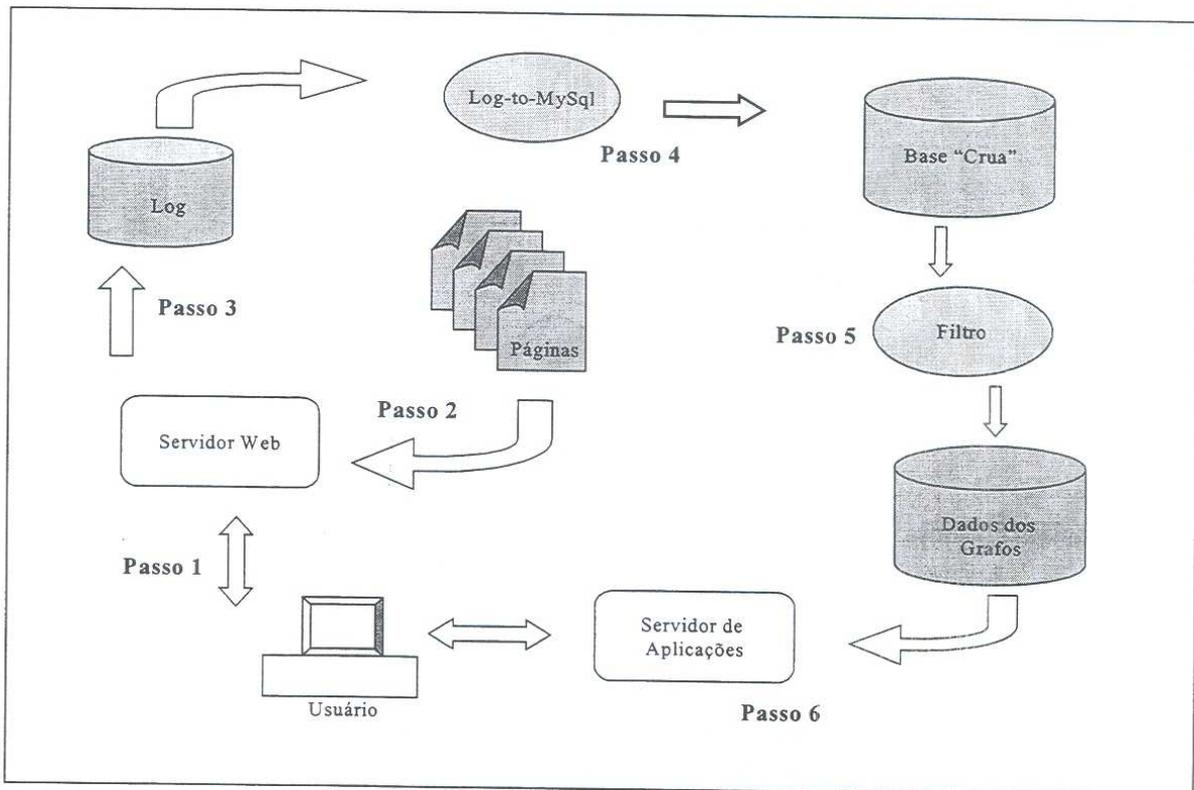
Tabela	Atributos
Usuário	Id_usuário, id_grupo
Local	Id_local, nome
Grupo	Id_grupo, nome, id_local
Acesso	Id_usuário, pag_referente, pag_acessada, n_ocorrencias

*Tabela 5.3 – Principais tabelas*

A tabela ACESSO será utilizada para representar as arestas do grafo. Considerando uma aresta  $\{p1,p2\}$ , iremos considerar que  $p2$  será a página acessada e  $p1$  a página que faz a referência a  $p2$ . O peso que esta aresta irá receber será o número de ocorrências deste acesso.

### 5.6.2. Esquema Geral

Neste tópico apresenta-se o esquema geral que representa a interação da ferramenta de suporte à navegação com o servidor web e o usuário.



**Figura 5.4 – Esquema Geral**

O usuário interage com o Portal navegando pelas páginas (passo1). A medida que o usuário navega pelas páginas acessadas pelo servidor web (passo 2) suas preferências são registradas nos registros de log (passo 3) compondo. Os registros da log do servidor web são lidos por um programa e gravados em uma base de dados relacional (passo 4) de forma a possibilitar uma manipulação mais ágil (dado estruturado).

Para poder descobrir as arestas do grafo de acesso, os dados são tratados por um outro filtro que gera então a base final com todos os dados necessários para descobrir o perfil de

cada grupo (passo 5). Este perfil é utilizado por um outro aplicativo que detecta as páginas acessadas pelo usuário durante a navegação e sugere páginas relacionadas baseado no comportamento de todo o grupo (passo 6).

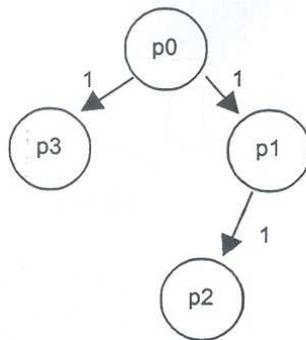
### 5.6.3. Processamento dos Grafos

Este tópico apresenta as regras empregadas para o processamento dos grafos obtidos a partir da análise da log do servidor web.

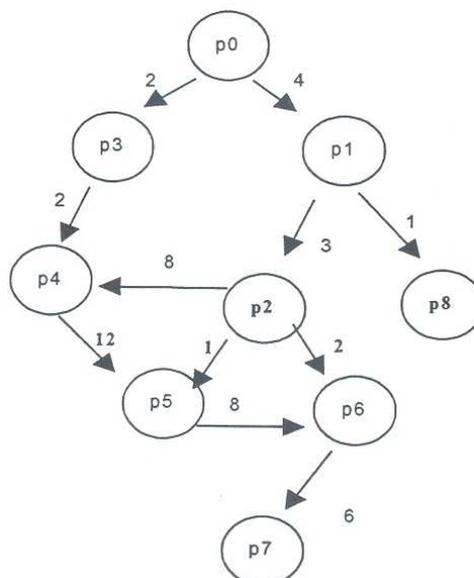
1. Inicialmente o grafo de cada grupo possuirá apenas a página inicial:



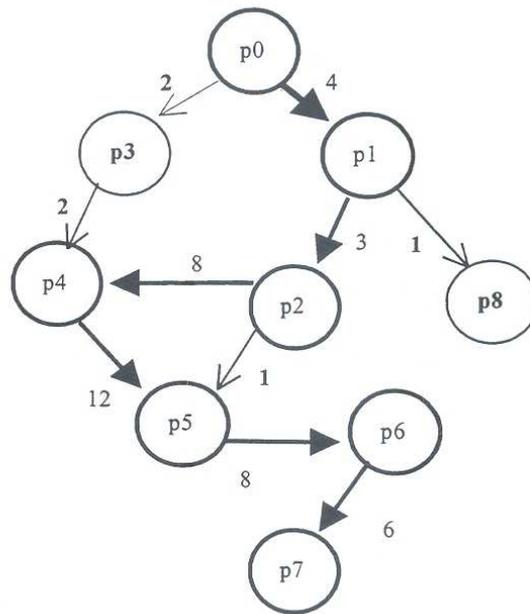
2. A partir da log de acesso, outros vértices são descobertos e as arestas são montadas, com peso 1.



3. Cada aresta terá seu peso incrementado a cada visita recebida. (Ou seja, os usuários percorrendo o mesmo caminho várias vezes).



4. Depois de um determinado período podemos já traçar o perfil do grupo, como um passeio no grafo cujas arestas possuem o maior peso, conforme visto na figura abaixo em **negrito**:



#### 5.6.4. Implementação Computacional

A ferramenta de suporte à navegação em Portal Corporativo foi desenvolvida utilizando os seguintes componentes tecnológicos:

- Máquina PC utilizando sistema operacional Linux
- WebServer Apache 1.3
- Application Server Tomcat 3.2
- Base de Dados MySQL 3.22
- Java 2 EE SDK
- Perl 5.

A escolha destas ferramentas foi baseada principalmente no fato de sua maioria se enquadrar na definição de software *Open Source*, ou seja, suas licenças não exigem nenhum tipo de ônus por utilização do software nem por utilização do código fonte para qualquer fim, seja para simples leitura, modificação ou redistribuição. A única exceção a esta regra é a tecnologia Java, pertencente a Sun Microsystems, mas mesmo assim esta permite a livre

utilização de seus produtos para fins comerciais e disponibiliza o código fonte para pesquisa e desenvolvimento.

Esta característica, além de tornar estas ferramentas mais robustas, permitiu também que fossem portadas para vários ambientes. Todas as ferramentas podem ser instaladas na grande maioria dos sistemas UNIX e também em plataforma Win32. O grande destaque neste quesito é a tecnologia Java, que foi projetada para poder ser executada em qualquer arquitetura.

O servidor Apache foi escolhido também [APA01] por ser um dos servidores Web mais utilizados segundo pesquisa da NetCraft, chegando a ser usada por 60% dos sites. O servidor de aplicações Tomcat foi desenvolvido pela mesma fundação e implementa a especificação 2.2 de Servlets. O protótipo implementado, ainda não foi otimizado para obter um desempenho melhor, mas é o ideal para um ambiente de desenvolvimento e pode ser substituído em um ambiente de produção por qualquer outro servidor que implemente a mesma especificação de Servlets.

Outra tecnologia envolvida foi Perl, que é uma linguagem de script também portada para vários sistemas operacionais. Possui como grande característica a habilidade de gerenciar arquivos textos de maneira simples, eficiente e prática devido seu grande suporte a expressões regulares. Isto o torna ideal para interpretar arquivos de log, e é para isto que ele é utilizado neste projeto, interpretando a log do servidor web e gravando os seus dados em uma base de dados. A base de dados utilizada é o MySQL, que é o servidor de banco de dados *Open Source* mais utilizado. Entretanto, pode ser substituído por qualquer servidor que suporte SQL ANSI.

#### **5.6.5. Especificação dos Processos**

A ferramenta é composta por três processos: dois são responsáveis pela extração dos dados, executados em *batch* e o outro é utilizado para interpretar o grafo gerado, interagindo com o usuário.

São dois os programas responsáveis pela extração dos dados. Eles podem ser executados em períodos determinados (por exemplo, uma vez ao dia, à meia-noite). O

primeiro programa utilizado, chamado *apachedb* escrito em Perl, recebe como parâmetro a log de acesso do servidor Apache. Este programa então interpreta cada linha da log e alimenta uma base de dados MySQL devidamente normalizada. Este programa foi modificado apenas para suportar SQL ANSI, não utilizando as extensões suportadas pelo MySQL.

O programa possui algumas características que não são exploradas por esta ferramenta. Três tipos de logs são reconhecidos: o log padrão, o log estendido, que possui algumas informações complementares, e o log que facilita o rastreamento do usuário. Este último é gerado por um módulo adicional do apache, chamado *modusertrack* que informa a página que referenciou o acesso e o nome do cookie que foi enviado ao usuário para futuros acessos. O formato de log utilizado foi o log padrão, pois assim a portabilidade do sistema para outros servidores web se torna muito maior, não dependendo de módulos especiais do Apache.

Outra característica seria a possibilidade de gerar os dados em tempo real, utilizando a capacidade do Apache de emitir a log diretamente para um programa. Esta possibilidade não foi explorada pois a ferramenta é dependente de um outro programa que irá gerar os dados sobre os grafos. Este programa foi escrito em Java e acessa a base gerada pelo programa *apachedb*, interpretando-a e descobrindo qual o provável caminho que cada usuário tomou, gerando então os dados para o grafo de acesso.

Para decidir sobre as páginas que referenciaram cada acesso, o programa lê cada registro realizado e descobre qual foi o acesso imediatamente anterior efetivado pelo mesmo usuário, caso a diferença no tempo de acesso de cada uma não seja maior que um período determinado  $t$ , assume-se que esta página referenciou a primeira. Este período  $t$  será denominado de “período de inatividade durante uma determinada navegação em um Portal” ou simplesmente de “período de inatividade”.

O servlet de referência é o programa que interage com o usuário, ele interpreta o grafo gerado pelos dois programas anteriores e aparece como um *frame* na navegação do usuário, sugerindo *links* baseado na página que o usuário está visualizando no momento. Caso este troque a página através de algum *link*, o servlet atualiza seu estado automaticamente.

O funcionamento do servlet é simples. Primeiramente ele detecta a página que o usuário está acessando no momento e realiza uma pesquisa no grafo para determinar qual vértice esta página representa. Uma vez descoberto este vértice, verifica-se quais são as arestas que partem deste grafo que possuem o maior peso. As três primeiras ocorrências, se existirem, serão exibidas na tela. Para descobrir a troca de páginas o servlet gera algumas instruções JavaScript programando o evento de mudança de páginas (chamado *onUnload*) do frame de navegação correspondente para que este notifique o servlet.

#### 5.6.6. Determinação do Período de Inatividade Máximo

O período de inatividade máximo  $\Delta_{\max}$  é calculado para cada usuário baseado na média de uma amostra de seu comportamento dos últimos acessos realizados. Este cálculo é executado juntamente com o processo de geração dos dados para os grafos, logo esta variação de tempo pode variar de acordo com cada usuário e da variação do comportamento do mesmo.

Detalhando esta característica tem-se que, dado que um usuário acesse as páginas A, B e C nesta sequência, e que os tempos de acesso de AB e BC sejam respectivamente 1 min e 3 min, calcula-se o tempo  $\Delta_{\max}$  será da ordem de 2 minutos em média.

A determinação do tempo médio de acesso por usuário é crucial para o processo de análise da log do servidor web uma vez que o padrão de log nos servidores web apresenta apenas o horário da consulta, não considerando o tempo de intervalo entre duas páginas. Com base no cálculo da amostra, é possível especificar o comportamento padrão de um determinado usuário e, assim, analisar se o registro de acesso a uma página efetivado na log diz respeito a continuação da navegação ou trata-se de nova trilha que o usuário está percorrendo.

Matematicamente, o cálculo do tempo médio de acesso entre páginas -  $\Delta_{\max}$  - pode ser definido como:

$$\Delta_{\max} = [ \sum(a_1, a_n) ] / n$$

onde

$a_1$  - primeiro elemento da amostra (diferença de horário entre duas páginas)

$a_n$  - n-ésimo elemento da amostra

$n$  - número de elementos da amostra

A diferença de horário entre duas páginas será equivalente ao ramo do grafo que está sendo analisado para efeito de identificação do perfil de um usuário. Exemplificando, caso o usuário tenha acessado a página A em 10:00:00 e a página B em 10:03:45, a diferença de horário entre A e B será de 3 minutos e 45 segundos (igual a  $a_1$ ). Ainda, se o acesso para a página C ocorreu em 10:05:23, a diferença de horário entre B e C será de 1 minuto e 38 segundos (igual a  $a_n$ ). Considerando a fórmula, tem-se que o somatório de  $a_1$  até  $a_n$  será de 5 minutos e 23 segundos para 2 ramos do grafo ( $n = 2$ ). Assim, o tempo médio de acesso entre páginas -  $t$  - é calculado em 2 minutos e 41.5 segundos.

Pela análise deste comportamento percebe-se que é possível considerar tão somente os intervalos externos da trilha definida no grafo pelo comportamento de navegação do usuário. Assim, a possível simplificação é sugerida:

$$\Delta_{\max} = [a_n - a_1] / n$$

Esta simplificação encerra um problema de determinar qual é o intervalo máximo que pode ser considerado entre o final de uma trilha e o início de outra. A proposta é que tal número referencial, vamos denominar de unidade básica de referência de navegação ( $U_n$ ) no site, seja obtida a partir de uma primeira experimentação conduzida de forma dirigida para o estabelecimento deste parâmetro - o apêndice A - Unidade Básica de Referência de Navegação - detalha esta discussão. Tal avaliação prévia permitirá definir qual unidade básica de referência de navegação ( $U_n$ ) servirá de referência para filtrar a amostragem. Assim, somente serão considerados elementos da amostra aqueles elementos  $a_n$  que forem resultado de uma diferença de horário entre páginas que seja menor ou igual a unidade básica de referência de navegação ( $U_n$ ).

### 5.6.7. Limitações

Existem algumas limitações na abordagem utilizada cuja solução está além do escopo deste projeto. As únicas páginas consideradas neste sistema são páginas estáticas. Qualquer acesso a programas (como CGI's ou servlets) não estão sendo considerados.

A identificação de cada usuário é feita simplesmente através do seu número IP, ou seja, assume-se que o usuário utilize sempre o mesmo endereço IP. Outros métodos de identificação podem ser utilizados para que o usuário possa realizar seu acesso por várias máquinas ou através de vários endereços IPs diferentes, mas isto exigiria um mecanismo de autenticação mais complexo.

Não é garantido que todo acesso a página por um usuário gere uma ocorrência na log de acesso. Com isto algumas arestas no grafo podem deixar de serem contabilizadas. O cliente (navegador) pode estar utilizando uma *cache* para as páginas utilizadas mais frequentemente e não consultar o webservice para versões mais recentes.

Isto dependerá da configuração do cliente e seu comportamento não pode ser garantido. Uma possível solução para isto seria acrescentar instruções nas páginas utilizadas para que o *cache* não seja utilizado, esta configuração pode ser feita também diretamente no servidor web. Entretanto, em ambos os casos deve-se levar em conta o tráfego de dados adicional na rede.

O problema da falta da garantia do que ocorre no cliente também acontece com o servlet, que necessita que o cliente execute algumas instruções JavaScript para que o servlet seja notificado da mudança de páginas. O cliente pode desativar ou não suportar estes comandos e não há como identificar esta ocorrência, simplesmente assume-se que o cliente estará com JavaScript habilitado.

## 5.7. Resultados Obtidos

O processo de suporte a navegação está em fase de testes no laboratório da Intranet do HSBC Bank Brasil S.A. Para a efetivação destes testes considerou-se que os acessos eram efetivados por usuários de um mesmo grupo (funcionários pertencentes a um mesmo departamento). O foco dos testes está centrado no conjunto de normas que atualmente suportam as atividades da instituição, sejam as atividades de retaguarda (*back-office*) ou o atendimento das agências (*front-office*). Tais normas, também denominadas normativos, constituem um conjunto significativo de informações organizadas como páginas da Web num total de 221 documentos. O volume de dados relativo aos normativos é da ordem de **84.948**

**Kbytes** distribuídos em **2.551 arquivos**. A média aproximada de 33 Kb por arquivo deve-se a uma recomendação do departamento de Telecomunicações (IT/Telecom) que aponta o tamanho máximo de 40 Kb por página, devido aos pontos remotos da rede que operam com links de baixa velocidade e concorrem com outros serviços.

Este conjunto de normas está organizado de forma a contemplar de 3 a 4 níveis de navegação para atingir o detalhe do normativo. Assim, o usuário acessa uma página principal e precisa navegar por outras três páginas, em média, antes de atingir seu objetivo. Com o uso da ferramenta de suporte à navegação, links relacionados são sugeridos conforme o peso atribuído, o que deve representar a redução no tempo total de acesso.

Visando avaliar o ganho obtido com o emprego da ferramenta apresentada foram consideradas 5 normas padrões da empresa e realizadas 5 medições para cada navegação em dois modos. O primeiro realizando uma navegação convencional onde o usuário deve ler as páginas acessadas e decidir pelo detalhe para prosseguir na navegação conforme o contexto. Neste caso o usuário tende a percorrer todo o caminho oferecido. Este primeiro modo de navegação será referenciado como Padrão de Referência. O segundo modo monitorado deverá contemplar o uso da ferramenta de suporte à navegação consistindo no teste proposto. Na tabela a seguir são apresentados os resultados obtidos com a navegação convencional obtida onde cada normativo equivale a uma rota de navegação.

Rota	Pag	Página	Bytes	1	2	3	4	5
1	1	..creditoimobiliario base.html	40.568	0:00:43	0:00:33	0:00:39	0:00:35	0:00:45
1	2	..creditoimobiliario info ger1.html	9.025	0:00:10	0:00:07	0:00:09	0:00:08	0:00:10
1	3	..avaliacoes imobiliarias base.html	7.382	0:00:08	0:00:06	0:00:07	0:00:06	0:00:08
1	4	..avaliacoes imobiliarias proc op1.html	9.787	0:00:10	0:00:08	0:00:09	0:00:09	0:00:11
1	5	..avaliacoes imobiliarias proc op2.html	8.005	0:00:09	0:00:07	0:00:08	0:00:07	0:00:09
2	1	..leasing conveniosvendedor base.html	23.109	0:00:25	0:00:19	0:00:22	0:00:20	0:00:25
2	2	..leasing conveniosvendedor info ger.html	8.491	0:00:09	0:00:07	0:00:08	0:00:07	0:00:09
2	3	..leasing conveniosvendedor base.html	37.079	0:00:40	0:00:31	0:00:36	0:00:32	0:00:41
2	4	..leasing conveniosvendedor proc op.html	7.938	0:00:08	0:00:07	0:00:08	0:00:07	0:00:09
2	5	..leasing conveniosvendedor consultas proc op.html	38.082	0:00:41	0:00:31	0:00:37	0:00:33	0:00:42
3	1	..hsbcpremier base.html	6.580	0:00:07	0:00:05	0:00:06	0:00:06	0:00:07
3	2	..hsbcpremier info ger.html	9.025	0:00:10	0:00:07	0:00:09	0:00:08	0:00:10
3	3	..hsbcpremier proc op.html	53.815	0:00:57	0:00:44	0:00:52	0:00:47	0:00:59
3	4	..hsbcpremier modelotermodeopcao.html	43.233	0:00:46	0:00:36	0:00:42	0:00:38	0:00:47
4	1	..cctdepartamentos base.html	26.197	0:00:28	0:00:22	0:00:25	0:00:23	0:00:29
4	2	..cctdepartamentos info ger.html	19.305	0:00:21	0:00:16	0:00:19	0:00:17	0:00:21
4	3	..cctdepartamentos proc op.html	6.958	0:00:07	0:00:06	0:00:07	0:00:06	0:00:08
5	1	..regionais/bhsi life protection/lifeprotection base.html	4.094	0:00:04	0:00:03	0:00:04	0:00:04	0:00:04
5	2	..regionais/bhsi life protection/lifeprotection info ger.html	13.937	0:00:15	0:00:11	0:00:13	0:00:12	0:00:15
5	3	..regionais/bhsi life protection/lifeprotection proc op.html	3.777	0:00:04	0:00:03	0:00:04	0:00:03	0:00:04
5	4	..regionais/bhsi life protection/lifeprotection historico.html	12.744	0:00:14	0:00:11	0:00:12	0:00:11	0:00:14

**Tabela 5.4 – Resultados de navegação – Modo Convencional**

Considerando, como exemplo, que o usuário deseja acessar informações relativas ao Leasing deverá acessar as páginas apresentadas na rota 2. Assim, considerando um comportamento padrão de navegação, deverão ser acessadas 5 páginas até atingir a informação desejada (rota 2 página 5). Simplificando estes valores pode-se consolidar a tabela conforme segue:

Rota	Normativo	1	2	3	4	5	Média
1	Crédito Imobiliário	0:01:20	0:01:02	0:01:12	0:01:05	0:01:22	0:01:12
2	Leasing Convênios Vendor	0:02:02	0:01:35	0:01:50	0:01:40	0:02:06	0:01:51
3	HSBC Premier	0:02:00	0:01:33	0:01:48	0:01:39	0:02:04	0:01:49
4	Controle de Contas Transitórias	0:00:56	0:00:43	0:00:50	0:00:46	0:00:58	0:00:51
5	Life Protection	0:00:37	0:00:28	0:00:33	0:00:30	0:00:38	0:00:33

*Tabela 5.5 – Consolidado navegação – Modo Convencional*

Na tabela a seguir são apresentados os valores de navegação obtidos conforme o segundo modo que contempla o uso da ferramenta de suporte à navegação.

Rota	Pag	Página	Bytes	1	2	3	4	5
1	1	..creditoimobiliario base.html	40.568	0:00:15	0:00:12	0:00:13	0:00:12	0:00:14
1	5	..avaliacoes imobiliarias proc op2.html	8.005	0:00:09	0:00:07	0:00:08	0:00:07	0:00:09
2	1	..leasing conveniosvendor base.html	23.109	0:00:14	0:00:11	0:00:12	0:00:11	0:00:14
2	5	..leasing conveniosvendor consultas proc op.html	38.082	0:00:17	0:00:09	0:00:16	0:00:12	0:00:14
3	1	..hsbcpremier base.html	6.580	0:00:07	0:00:05	0:00:06	0:00:06	0:00:07
3	4	..hsbcpremier modelotermodeopcao.html	43.233	0:00:46	0:00:36	0:00:42	0:00:38	0:00:47
4	1	..cctdepartamentos base.html	26.197	0:00:28	0:00:22	0:00:25	0:00:23	0:00:29
4	3	..cctdepartamentos_proc op.html	6.958	0:00:08	0:00:06	0:00:07	0:00:10	0:00:07
5	1	..regionais/bhsi life protection/lifeprotection base.html	4.094	0:00:04	0:00:03	0:00:04	0:00:04	0:00:04
5	4	..regionais/bhsi life protection/lifeprotection historico.html	12.744	0:00:15	0:00:12	0:00:13	0:00:12	0:00:14

*Tabela 5.6 – Resultados de navegação – Modo apoiado Ferramenta*

Consolidando estes dados no mesmo padrão adotado para a amostra de Referência tem-se:

Rota	Normativo	1	2	3	4	5	Média
1	Crédito Imobiliário	0:00:24	0:00:19	0:00:21	0:00:19	0:00:23	0:00:21
2	Leasing Convênios Vendor	0:00:31	0:00:20	0:00:28	0:00:23	0:00:28	0:00:26
3	HSBC Premier	0:00:53	0:00:41	0:00:48	0:00:44	0:00:55	0:00:48
4	Controle de Contas Transitórias	0:00:36	0:00:27	0:00:32	0:00:33	0:00:36	0:00:33
5	Life Protection	0:00:19	0:00:15	0:00:17	0:00:16	0:00:18	0:00:17

*Tabela 5.7 – Consolidado navegação – Modo apoiado Ferramenta*

Ao comparar os tempos obtidos pela amostra de referência e pela amostra de teste obtêm-se os seguintes dados:

Rota	Normativo	Referência	Teste	Ganho
1	Crédito Imobiliário	0:01:12	0:00:21	71,0%
2	Leasing Convênios Vendor	0:01:51	0:00:26	76,6%
3	HSBC Premier	0:01:49	0:00:48	55,8%
4	Controle de Contas Transitórias	0:00:51	0:00:33	35,4%
5	Life Protection	0:00:33	0:00:17	48,6%

*Tabela 5.8 – Resultados Alcançados*

Tais valores apontam para um potencial ganho de produtividade dos funcionários na pesquisa de documentos que sejam suportados por esta ferramenta. Observa-se que os ganhos são maiores quanto maior o número de páginas e a extensão das mesmas. Numa projeção conservadora pode-se considerar um incremento de produtividade da ordem de 30% que sugere uma vantagem competitiva em relação a outras ferramentas de apoio a pesquisa.

## 5.8. Conclusão

Este capítulo apresenta a proposta de implementação de ferramenta de suporte à navegação baseado no perfil do usuário denotando seu uso em Portais Corporativos, de uma forma geral, e notadamente em Portais de empresas financeiras. São detalhados os requisitos a serem contemplados por esta ferramenta, bem como os requisitos que não serão contemplados. Tais relações de requisitos baseiam-se na proposta de Eckerson [ECK01].

Faz-se uma apresentação dos resultados esperados pelo modelo proposto, como a definição de tecnologia, o suporte a decisão, a predictibilidade e a personalização. Na sequência apresenta-se uma proposta de modelagem para a solução contemplando-se os passos para a obtenção da relação de páginas candidatas a apresentação. Detalha-se a solução do ponto de vista computacional contemplando a descrição do mapeamento da navegação em um grafo, discutindo o problema da determinação do tempo de inatividade. Ao final são referenciadas algumas limitações desta abordagem, bem como são apresentados os resultados obtidos, denotando os possíveis ganhos com o emprego desta proposta. O próximo tópico apresenta uma avaliação final deste trabalho.

## CONCLUSÃO

Os Portais Corporativos, tendo como “ancestrais” os sistemas de suporte à decisão e os sistemas de informações gerenciais, são o próximo passo no desenvolvimento de interfaces de usuário às informações corporativas. Adaptando o ambiente organizacional para atender às necessidades dos usuários e otimizar a interação, a distribuição e a gerência dos recursos informacionais internos e externos, o portal corporativo permite que os usuários acessem as informações corporativas de forma mais ágil e personalizada, resultando, teoricamente, em aumento de produtividade, redução de custos e aumento de competitividade da organização.

Sua capacidade de facilitar o acesso dos usuários às informações institucionais está intrinsecamente relacionada à facilidade de uso, aprendizado e satisfação do usuário com sua interface web. Para conseguir concretizar esse benefício, portanto, é fundamental que o projeto do portal corporativo leve em consideração a interação dos usuários com sua interface e tenha como objetivo minimizar seus problemas de usabilidade.

A proposta de uma ferramenta de suporte ao acesso às informações disponibilizadas num Portal Corporativo, baseada na avaliação do perfil de navegação do usuário (característica de usabilidade) será útil para qualquer instituição que mantenha informações sob a forma digital. Este trabalho mostra que é possível rastrear o comportamento de usuários dentro de um site de uma Intranet, notadamente em Portais Corporativos. Este comportamento, ou perfil, pode ser modelado e armazenado em uma base de conhecimento que pode ser utilizado para os mais diversos fins.

Neste projeto, foi utilizado como exemplo um *servlet* que mostra sugestões de páginas para o usuário durante a navegação tendo como base o perfil de todo o grupo o qual este

usuário pertence. Tais sugestões irão caracterizar um ganho de produtividade pela diminuição no tempo de busca de informações.

Entretanto, para modelar o comportamento foi necessário assumir várias premissas, e algumas heurísticas foram utilizadas para determinar o relacionamento entre duas páginas. Ainda assim, o problema em sua essência é resolvido e utilizando alguns mecanismos mais complexos pode-se dispensar algumas destas premissas e determinar funções mais exatas para determinar estes relacionamentos, que poderá ser o objeto de trabalhos futuros.

Dois itens adicionais são importantes na conclusão deste trabalho: a) Resultados Indiretos e b) Pesquisas Complementares.

#### **a) Resultados Indiretos**

Em função da análise realizada visando conceituar os requisitos da ferramenta foram obtidos alguns resultados indiretos que merecem registro. A partir das diversas características encontradas e a diversidade de conceitos empregados, foi necessário estabelecer uma forma padronizada de tipificar os Portais.

Foi evidenciado que as funções mais importantes de um portal corporativo são: gestão do conhecimento, suporte à decisão e processamento cooperativo. Conforme apresentado, o uso de certas nomenclaturas pode levar a interpretações equivocadas sobre as reais características de um Portal. Baseado na avaliação da principal utilização do Portal foi sugerida uma tipificação representada no quadro *Tipos de Portais baseado na utilização*.

#### **b) Pesquisas Complementares**

A partir do trabalho realizado algumas pesquisas derivadas foram identificadas porém não chegaram a ser realizadas no escopo deste trabalho. Pode-se sugerir como itens de pesquisa complementar:

- **Avaliação de Interfaces externas** – possibilitar a integração da ferramenta com produtos de mercado e aplicações customizadas, definindo formatos que

possibilitem apresentar as sugestões de navegação obtidas com a ferramenta de suporte à navegação.

- **Interfaces programáveis** - definir interfaces públicas (API - *Application Programming Interface*) que possibilitem programar a ferramenta a partir de outros aplicativos.
- **Determinação automática do período de inatividade** – como forma de incrementar a funcionalidade da ferramenta verifica-se a possibilidade de determinar de forma automática o período de inatividade a partir do emprego de agentes que, dinamicamente, avaliassem o tempo de inatividade viabilizando uma maior interação da ferramenta com o usuário.
- **Classificação dos usuários de acordo com o perfil** – é possível perceber o potencial de classificar os usuários em grupos não em função de sua localização geográfica (fundamento deste estudo realizado) mas sim de seu perfil de navegação. Neste caso a análise de grafos similares levaria o usuário a ser escalado para determinado grupo.

## Referências Bibliográficas

- [ALA00] \_\_\_\_\_. *Access Log Analyzers*. Abril, 2000. [<http://www.uu.se/Software/Analyzers/Access-analyzers.html>]
- [APA01] \_\_\_\_\_. *The Apache Software Foundation*. Novembro, 2001 - [<http://www.apache.org>]
- [BAE99] BAEZA-YATES, RICARDO. & RIBEIRO-NETO, BERTHIER. *Modern Information Retrieval*. ACM Press, 1999.
- [BKA97] KAHLE, B. *Archiving the Internet*. Abril, 2000. - [[http://www.alex.com/~brewster/essays/sciam\\_article.html](http://www.alex.com/~brewster/essays/sciam_article.html)]
- [BRI00] \_\_\_\_\_. *How to choose an enterprise information portal solution: Seybold's 10 keys to enterprise portals*. Brio Technology. Palo Alto, CA. Agosto, 2000. [[http://www.brio.com/library/white\\_papers/pdf\\_files/BrioPortal\\_Seybold.pdf](http://www.brio.com/library/white_papers/pdf_files/BrioPortal_Seybold.pdf)].
- [BRO01] \_\_\_\_\_. *BroadVision One-To-One Content*. Dezembro, 2001. [<http://www.broadvision.com/content/products/datasheets/Content.pdf>]
- [COL99] COLLINS, DANIEL. *Data warehouses, enterprise information portal, and the SmartMart meta directory*. Information Builders Systems Journal, v. 12, n. 2, p. 53-61, Mar./Apr. 1999.
- [COO97] R.COOLEY, B.MOBASHER E J.SRIVASTAVA. *Web Mining: Information and Pattern Discovery on the World Wide Web*. In: 9th IEEE International Conference on Tool with Artificial Intelligence. Newport Beach, 1997. Anais. IEEE, 1997, pp 558-567.

- [COU94] COULORIS, GEORGE. JEAN DOLLIMORE. TIM KINDBERG. *Distributed Systems*. Addison-Wesley, 1996.
- [DAI01] \_\_\_\_\_. *Enterprise information portal requirements*. Decision processing brief, DP-99-02. Morgan Hill, CA: Database Associates International, Jan. 1999. Abril 2001. [<http://www.decisionprocessing.com/papers/eip2.doc>].
- [DAI99] \_\_\_\_\_. *The enterprise information portal marketplace*. Decision processing brief, DP-99-01. Morgan Hill, CA: Database Associates International, Jan. 1999. Setembro, 2001. [<http://www.decisionprocessing.com/papers/eip1.doc>].
- [DAV01] DAVYDOV, MARK M. *EIP: the second wave*. Intelligent Enterprise, v. 3, n. 4, Mar. 2000. Junho, 2001. [<http://www.intelligententerprise.com/000301/supplychain.shtml>].
- [DIG01] JAMES DIGGES. *Financial services industry - Achieving Value from Knowledge Management*. Novembro, 2001 [<http://www.kmmagazine.com/Downloads/KMFinancialSector2001.pdf>]
- [ECK01] ECKERSON, WAYNE. *15 rules for enterprise portals*. Oracle Magazine, v. 13, n. 4, p. 13-14, July/Aug. 1999. Agosto, 2001. [<http://www.oracle.com/oramag/oracle/99-Jul/index.html?49ind.html>].
- [EIP00] \_\_\_\_\_. *Defining the enterprise information portal*. July 1999. Abril, 2001. [<http://www.dkms.com/EIPDEF.html>].
- [ETZ96] ETZIONI, O. *The world wide web: Quagmire or gold mine*. Communications of the ACM, 39( 11): 65- 68. Junho, 2000.
- [FAY96] U.M.FAYYAD, G.PIATETSKY-SHAPIRO E P.SMYTH. *From Data Mining to Knowledge Discovery: An Overview*. In: Advances in Knowledge Discovery and Data Mining. Menlo Park, AAAI Press, 1996, pp 11-34.
- [FRA01] CARL FRAPPAOLO. *Economy Favors Expansion of Content and Knowledge Management Resources*, Delphi Group. Novembro, 2001. [[http://www.delphigroup.com/about/pressreleases/more\\_press\\_releases.htm?narrow=KM](http://www.delphigroup.com/about/pressreleases/more_press_releases.htm?narrow=KM)]
- [GOG01] IBRAHIM GOGUS. *Knowledge management systems*. Novembro, 2001.[[http://www.knowledge-management.co.uk/KM\\_FAQS.asp](http://www.knowledge-management.co.uk/KM_FAQS.asp)]
- [INM00] WILLIAM H. INMON. *An architecture for managing click stream data*. Abril, 2001 [[http://www.billinmon.com/library/library\\_frame.html](http://www.billinmon.com/library/library_frame.html)]

- [INW01] \_\_\_\_\_. *Gartner Group apresenta novas tendências para o mercado de TI*. Dezembro, 2001. [<http://www.informationweek.com.br/psi/artigo.asp?id=11971>]
- [ISO91] \_\_\_\_\_. *Information technology - software product evaluation: quality characteristics and guidelines for their use*. ISO/IEC 9126. 1991.
- [ISO98] \_\_\_\_\_. *Part 11 Ergonomic requirements for office work with visual display terminals: Part 11 -Guidelines on usability*. ISO 9241. 1998.
- [KDD01] \_\_\_\_\_. *WebKDD - Workshops in Web Mining*. Novembro, 2001. [<http://robotics.Stanford.edu/~ronnyk/webkdd2000/index.html>]
- [KIM00] RALPH KIMBALL E RICHARD MERZ. *The Data Webhouse Toolkit - Building the Web - Enabled Data Warehousing* John Willey & Sons - 2000.
- [KST01] \_\_\_\_\_. *K-station - Lotus knowledge management portal*. Dezembro, 2001. [<http://www.lotus.com/>]
- [MLL01] \_\_\_\_\_. *Personalized access to critical information and resources*. Dezembro, 2001. [<http://www.opentext.com/mylivelink/>]
- [MUR99] MURRAY, GERRY. *The portal is the desktop*. Intraspect, May/June 1999. Julho/2001. [[http://www.groupcomputing.com/Back\\_Issues/1999/MayJune1999/mayjune1999.html](http://www.groupcomputing.com/Back_Issues/1999/MayJune1999/mayjune1999.html)]
- [NAS00] O.NASRAOUI, H.FRIGOUI E A.JOSHI. *Mining Web Access Logs Using Relational Competitive Fuzzy Clustering*. Julho, 2001. [<http://csdeca.cs.missouri.edu/~joshi/web-mine/kdd.ps>].
- [PLU00] \_\_\_\_\_. *Plumtree corporate portal 3.0*. Abril, 2001. [<http://www.plumtree.com/products.html>].
- [PLU01] \_\_\_\_\_. *Plumtree: Internet Architecture*. Dezembro, 2001 [<http://www.plumtree.com/products/portal/architecture/>]
- [PSG00] \_\_\_\_\_. *Business portals: drivers, definitions, and rules*. Boston, MA: Patricia Seybold Group, 1999. Abril 2000. [<http://www.viador.com/pdfs/SeyboldWhitePaper.pdf>].
- [REY00] REYNOLDS, HADLEY & KOULOPOULOS, TOM. *Enterprise knowledge has a face*. *Intelligent Enterprise*, v. 2, n. 5, p. 29-34, Mar. 1999. Setembro, 2000. [<http://www.intelligententerprise.com/993003/feat1.shtml>]
- [SHI99] SHILAKES, CHRISTOPHER C. & TYLMAN, JULIE. *Enterprise information portals*. New York, NY: Merrill Lynch, 16 Nov. 1998. Janeiro, 2000. [<http://www.sagemaker.com/home.asp?id=500&file=Company/WhitePapers/>]

lynch.htm]

- [SIL96] A. SILBERSCHATZ E A.TUZHILIN. *What makes patterns interesting in knowledge discovery systems*. In: IEEE Transactions on Knowledge and Data Engineering. Vol. 8, No. 6, Dezembro 1996.
- [SKI01] DAVID J. SKYRME. *Measuring Intellectual Capital: A Plethora of Methods*. Business Intelligence (1998).
- [SPI00] M.SPILIOPOULOU E L.C.FAULSTICH. *WUM: A Tool for Web Utilization Analysis*. In: EDBT Workshop WEBDB'98. Valencia, Março 1998. Anais, 1998.
- [SPO00] \_\_\_\_\_. *Plumtree blossoms: new version fullfills enterprise portal requirements*. Boston, MA: Patricia Seybold Group, June 1999. Julho, 2000. [<http://www.plumtree.com/moreinfo/specialoffer.htm>].
- [STE02] THOMAS A. STEWART. *The Case Against Knowledge Management*. Business 2.0 Magazine. February 2002 Issue. Jan, 2002. [<http://www.business2.com/articles/mag/0,1640,36747,FF.html>]
- [VIA00] \_\_\_\_\_. *Enterprise information portals: realizing the vision of information at your fingertips*. Jan. 1999. Julho, 2001. [[http://www.viador.com/pdfs/EIP\\_white\\_paper\\_1\\_99.pdf](http://www.viador.com/pdfs/EIP_white_paper_1_99.pdf)].
- [W3C01] \_\_\_\_\_. *The World Wide Web Consortium (W3C)* – Novembro, 2001. [<http://www.w3.org/>]
- [WHI01] WHITE, COLIN. *Decision Threshold*. Intelligent Enterprise, v. 2, n. 16, p. 35-40, Nov. 1999. Setembro, 2001. [<http://www.intelligententerprise.com/991611/feat1.shtml>].
- [WSP01] \_\_\_\_\_. *WebSphere Personalization for Multiplatforms V4.0*. Dezembro, 2001. [<http://www-4.ibm.com/software/webservers/personalization/>]
- [ZAI98] O.R.ZAIANE, M.XIN E J.HAN. *Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs*. In: Advances in Digital Libraries. Santa Barbara, 1998.
- [ZUL01] FERNANDO ZULIANI. *WebSphere Personalization for Multiplatforms - The Personalization* Redbook-SG24-6214. Agosto, 2001. [<http://www.redbooks.ibm.com/>].

# Glossário

***Content Management*** - Gestão de Conteúdos - Um conjunto de tecnologias que visa tornar mais fácil a manutenção e administração de conteúdos, quer se apresentem em formato texto, gráficos ou dados.

***Customer Relationship Management (CRM)*** – Gestão de Relacionamento com o Cliente - tecnologia de suporte para lidar com clientes. Quando tais práticas são estendidas à *Internet*, são geralmente conhecidas por eCRM.

***Data Warehouse*** - Armazém de Dados - Tecnologia de infra-estrutura para desenvolver a análise estatística de largas quantidades de dados off-line (regra geral de clientes) arquivados. Pode indicar oportunidades úteis para vendas cruzadas.

***Enterprise Information Portal (EIP)*** – Portal de Informações Empresarias - Interface unificada para visualizar todas as aplicações, com poderosas funções de pesquisa e administração de conhecimento.

***Enterprise Resource Planning (ERP)*** – Planejamento de Recursos da Organização - Um sistema de software que tem módulos integrados para cada função da empresa, como, por exemplo, recursos humanos e departamento financeiro, possibilitando uma maior eficiência ao negócio, bem como a criação automática de relatórios detalhados de gestão e de previsões.

***Knowledge Management*** - Gestão do Conhecimento - A prática e as tecnologias relacionadas com a coleção, organização, partilha (ou disseminação) e análise do conhecimento da empresa. As tecnologias que facilitam esta administração incluem OCR (Optical Character Recognition), "pushing" e "data mining" (extração de dados).

***On Line Analytical Processing (OLAP)*** - Processamento Analítico Online - permite ao usuário extrair seletivamente e visualizar dados sob diferentes prismas.

**Customização** - A prática de "tailoring" (ajustar) o conteúdo da Web aos requisitos específicos de cada usuário em particular, que pode ser identificado pelo seu endereço de IP (*Internet Protocol*) ou pela presença de um "cookie" - uma pequena peça de script ou de software enviada para a máquina do usuário pelo servidor.

**XML Extensible Markup Language** - Subconjunto da linguagem SGML - Standard Generalized Markup Language, sistema padronizado de organização de documentos, desenvolvido pela International Organization for Standards (ISO). A XML foi projetada para facilitar a comunicação entre sistemas conectados na Internet, provendo uma maneira fácil de definir metadados associados ao conteúdo de recursos web.

**Firewall** - Dispositivos empregados na proteção de redes de computadores contra ataques externos.

# Apêndice A

## Unidade Básica de Referência de Navegação

Conforme visto neste trabalho, um dos pontos críticos para a análise dos registros de log do servidor web é a determinação do período máximo de inatividade  $\Delta_{\max}$ . Foi avaliado que o tempo de inatividade deve ser calculado em função de cada usuário. Tal definição deve-se ao fato de que há um tempo variável de leitura, avaliação e decisão para cada página pesquisada. De forma simplificada as atividades de leitura, avaliação e decisão de próxima página serão denominadas de navegação no site ou, apenas, navegação. Assim, o tempo de navegação de cada usuário é uma variável em função de seu padrão de acesso. Por consequência, o tempo de inatividade  $\Delta_{\max}$  será o valor superior ao tempo médio de navegação.

A determinação do tempo de navegação é uma função matemática bastante simples que reflete o tempo médio de acesso por usuário entre as páginas navegadas. A partir de uma amostra de registros de páginas acessadas, previamente filtradas para um determinado usuário, é possível especificar o comportamento padrão e, assim, avaliar se o registro de acesso a uma página efetivado na log diz respeito a continuação da navegação ou trata-se de nova trilha que o usuário está percorrendo.

A fórmula matemática do o cálculo do tempo médio de acesso entre páginas -  $\Delta_{\max}$  - pode ser apresentada como:

$$\Delta_{\max} = [ \Sigma(a_1, a_n) ] / n$$

onde

$a_1$  - primeiro elemento da amostra (diferença de horário entre duas páginas)

$a_n$  - n-ésimo elemento da amostra

$n$  - número de elementos da amostra

A diferença de horário entre duas páginas será equivalente ao ramo do grafo que está sendo analisado para efeito de identificação do perfil de um usuário. Exemplificando, caso o usuário tenha acessado a página A em 10:00:00 e a página B em 10:03:45, a diferença de horário entre A e B será de 3 minutos e 45 segundos (igual a  $a_1$ ). Ainda, se o acesso para a página C ocorreu em 10:05:23, a diferença de horário entre B e C será de 1 minuto e 38 segundos (igual a  $a_n$ ). Considerando a fórmula, tem-se que o somatório de  $a_1$  até  $a_n$  será de 5 minutos e 23 segundos para 2 ramos do grafo ( $n = 2$ ). Assim, o tempo médio de acesso entre páginas -  $\Delta_{\max}$  - é calculado em 2 minutos e 41.5 segundos.

Um dos aspectos problemáticos no cálculo deste tempo é que em Portais Corporativos onde existe um volume de acesso significativo, uma vez que o usuário transaciona constantemente com o site, pode ser inviável processar um conjunto elevado de registros do log do servidor web. O Portal Corporativo do HSBC, por exemplo, apresenta um volume mensal de 366.931 *pages view* apenas na *Home Page* - Fonte: *Web Trends - Dez/2001*. Assim, o tratamento dos registros da log do servidor web para filtrar uma amostra qualquer é extremamente caro. Uma forma de evitar todo este tratamento é trabalhar com o cálculo do tempo de navegação baseado nos intervalos externos da trilha.

Assim, a possível simplificação é sugerida:

$$\Delta_{\max} = [a_n - a_1] / n$$

Esta simplificação encerra um problema: Como determinar qual é o intervalo máximo que pode ser considerado entre o final de uma trilha e o início de outra uma vez que os valores intermediários são desprezados? A proposta é calcular um número referencial, vamos denominar de unidade básica de referência de navegação ( $U_n$ ) no site. A unidade de referência de navegação irá apontar o intervalo padrão de navegação no Portal. Este intervalo será considerado para efeito de tratamento dos registros da log, possibilitando contabilizar as páginas num intervalo de tempo que denota o comportamento padrão de navegação no site. Como exemplo, a unidade básica de referência do protótipo foi avaliada em 13:37 s.

Assim, dado este padrão, é possível selecionar amostras da log que estejam no limite deste intervalo. Tal avaliação prévia permitirá definir qual unidade básica de referência de navegação ( $U_n$ ) servirá para filtrar a amostragem. Assim, somente serão considerados elementos da amostra aqueles elementos  $a_n$  que forem resultado de uma diferença de horário entre páginas que seja menor ou igual a unidade básica de referência de navegação ( $U_n$ ).