

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIAS DA COMPUTAÇÃO – BACHARELADO

SISTEMA WEB PARA ADMINISTRAÇÃO DE IGREJAS
UTILIZANDO TÉCNICAS DE DATA MINING

LEMUEL SANTOS

BLUMENAU
2004

2004/2-31

LEMUEL SANTOS

SISTEMA WEB PARA ADMINISTRAÇÃO DE IGREJAS

UTILIZANDO TÉCNICAS DE DATA MINING

Trabalho de Conclusão de Curso submetido à Universidade Regional de Blumenau para a obtenção dos créditos na disciplina Trabalho de Conclusão de Curso II do curso de Ciência da Computação — Bacharelado.

Prof. Paulo Roberto Dias

**BLUMENAU
2004**

2004/2-31

SISTEMA WEB PARA ADMINISTRAÇÃO DE IGREJAS

UTILIZANDO TÉCNICAS DE DATA MINING

Por

LEMUEL SANTOS

Trabalho aprovado para obtenção dos créditos na disciplina de Trabalho de Conclusão de Curso II, pela banca examinadora formada por:

Presidente: _____
Prof. Paulo Roberto Dias – Orientador, FURB

Membro: _____
Prof. Maurício Capobianco Lopes, FURB

Membro: _____
Prof. Oscar Dalfovo, FURB

Blumenau, 08 de dezembro de 2004

Dedico especialmente aos meus queridos pais, Nilton e Regina pelo amor e carinho demonstrado a minha pessoa, e também a todos que de alguma forma contribuíram, incentivaram e apoiaram a idéia da realização deste trabalho.

Não diga que a vitória está perdida.
Tenha fé em Deus, tenha fé na vida.
Basta ser sincero e desejar profundo.
Você será capaz de sacudir o mundo.
Vai, tente outra vez.

Raul dos Santos Seixas

AGRADECIMENTOS

À Deus, o autor da vida.

Aos meus pais, Nilton e Regina que não mediram esforços para a realização deste sonho.

Aos meus amigos, Luciano Berto, Luciano Marquadt e Raphael Petters pelo incentivo e companheirismo durante estes anos de academia.

Ao meu orientador, Paulo Roberto Dias, por ter me ajudado em todos os momentos orientando-me neste projeto.

A todos os professores que ofereceram ajuda, apoio e ensinamentos a minha pessoa.

A Igreja Evangélica Assembléia de Deus de Blumenau que não mediu esforços para contribuir compartilhando os conhecimentos da administração interna da igreja e pelas horas e horas de teste neste sistema.

RESUMO

Este trabalho tem como principal objetivo auxiliar a administração de igrejas através de um sistema web, descentralizando os processos existentes entre uma igreja sede e suas congregações fazendo a atualização de dados e informações on-line. Objetiva-se também aplicar técnicas de *data mining* para prover relatórios para os administradores da igreja em suas tomadas de decisões. O sistema foi implementado em *Personal Home Page* (PHP) e banco de dados MySQL. O sistema já vem sendo testado e utilizado pela Igreja Evangélica Assembléia de Blumenau, Santa Catarina.

Palavras chaves: Sistemas; Igreja; Data Mining; WEB; PHP; MySQL.

ABSTRACT

This work has as main goal to help churches manager through a web system, not centralizing the process between a church and its congregation making the data and information's updates on-line. Another goal is to apply data mining techniques providing reports for the churches manager in their decisions. The system was implemented in Personal Home Page (PHP) and database MySQL. The system already has been tested and used by Assembly of God Gospel Church in Blumenau, Santa Catarina.

Key-Words: Systems; Church; Data Mining; Web; PHP; MySQL.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico do crescimento dos evangélicos no Brasil em relação a população.....	12
Figura 2 – Diagrama de atividade do caso de uso do desenvolvimento do software	23
Figura 3 – Diagrama de atividades do caso de uso iteração de software	25
Figura 4 – Processo de descoberta do conhecimento através do banco de dados	30
Figura 5 – Fórmulas para calcular entropia e ganho	36
Figura 6 – Visualização do processo de entradas	40
Figura 7 – Visualização do processo de saídas	40
Figura 8 – Visualização parcial da transação de informações na IEADB.....	42
Figura 9 – Visualização da listagem de membro no ROL-BL	44
Figura 10 – Visualização do cadastro de membro no ROL-BL	44
Figura 11 – Visualização dos dados pessoais do membro no GESIG.....	45
Figura 12 – Visualização dos usuários (administradores) envolvidos no sistema	47
Figura 13 – Diagrama de casos de uso do administrador central e do administrador global ...	48
Figura 14 – Diagrama de casos de uso do administrador congregacional.....	50
Figura 15 – Diagrama entidade relacionamento lógico.....	51
Figura 16 – Árvore de decisão gerada pelo algoritmo ID3 sem podagem de árvore	61
Figura 17 – Árvore de decisão gerada pelo algoritmo C4.5 com podagem de árvore	62
Figura 18 – Árvore construída pela ferramenta WEKA a partir do algoritmo id3.....	63
Figura 19 – Árvore construída pela ferramenta WEKA a partir do algoritmo c4.5	63
Figura 20 – Página inicial do sistema.....	65
Figura 21 – Menus de navegação para o nível 1 (administrador global)	66
Figura 22 – Menus de navegação para o nível 3 (administrador congregacional).....	67
Figura 23 – Visualização parcial do cadastro de membro do sistema.....	68
Figura 24 – Janela de visualização do membro da igreja.....	68
Figura 25 – Visualização do cadastro de entrada (nível 1 e 2 X nível 3).....	69
Figura 26 – Visualização da versão para impressão do recibo de entrada	69
Figura 27 – Visualização do cadastro de saída (nível 1 e 2 X nível 3)	70
Figura 28 – Visualização da versão para impressão do recibo de saída.....	70
Figura 29 – Visualização da listagem de cidade	71
Figura 30 – Visualização do relatório de membros.....	72
Figura 31 – Visualização da página de mineração de dados do sistema.....	75
Quadro 1 – Setores e congregações da IEADB.....	37
Quadro 2 – Dicionário de dados do sistema.....	55
Quadro 3 – Cálculo de entropia do atributo alvo	57
Quadro 4 – Cálculo de ganho do atributo número de membros da congregação.....	58
Quadro 5 – Cálculo de ganho do atributo número de contribuintes da congregação.....	59
Quadro 6 – Cálculo de ganho do atributo templo construído.....	60
Quadro 7 – Cálculo de ganho do atributo localização da congregação.....	60
Quadro 8 – Regras no sistema a partir da árvore de decisão gerada pelo algoritmo id3.....	73
Quadro 9 – Regras geradas a partir da árvore de decisão gerada pelo algoritmo c4.5.....	74

LISTA DE TABELAS

Tabela 1 – Tabela de decisão do andamento da congregação	56
--	----

LISTA DE SIGLAS

CGADB – Convenção Geral das Assembléias de Deus
CGI – *Common Gateway Interface*
DCU – Diagrama de Caso de Uso
DER – Diagrama Entidade Relacionamento
FTP – *File Transfer Protocol*
IBGE – Instituto Brasileiro de Geografia e Estatística
IEAD – Igreja Evangélica Assembléia de Deus
IEADB – Igreja Evangélica Assembléia de Deus de Blumenau
HTML – *Hyper Text Markup Language*
HTTP – *Hiper Text Transfer Protocol*
KDD – *Knowledge Discovery in Databases*
ODBC – *Open Database Connectivity*
PHP – *Personal Home Page*
SQL – *Structured Query Language*
TCO – *Total Cost of Ownership*
UML – Linguagem de Modelagem Unificada
URL – *Uniform Resource Locator*
XML - *eXtensible Markup Language*
W3C – *World Wide Web Consortium*
WEKA – *Waikato Environment for Knowledge Analysis*

SUMÁRIO

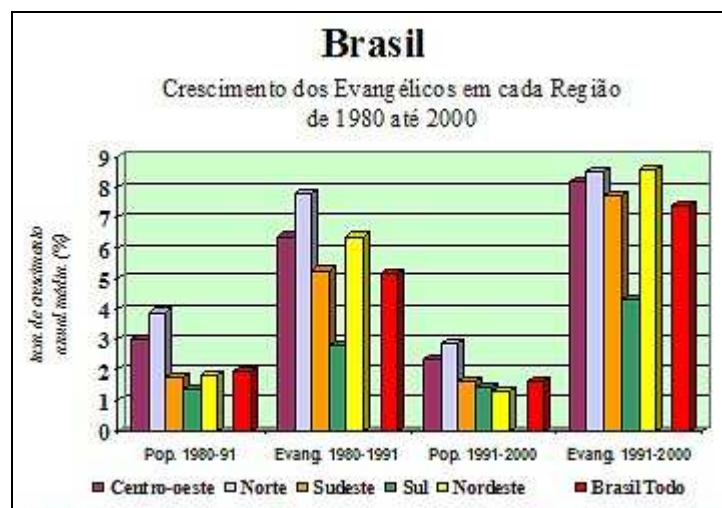
1 INTRODUÇÃO.....	12
1.1 OBJETIVOS.....	14
1.2 ESTRUTURA DO TRABALHO.....	14
2 SISTEMAS WEB.....	16
2.1 A WEB DINÂMICA.....	17
2.2 CARACTERÍSTICAS DO PROCESSO DE SISTEMAS WEB.....	18
2.3 PRINCIPAIS VANTAGENS DE SISTEMAS WEB.....	18
2.4 PRINCIPAIS DESVANTAGENS DE SISTEMAS WEB.....	19
2.4.1 SEGURANÇA.....	20
2.4.2 FALTA DE PADRÃO NA WEB.....	22
2.5 ATIVIDADES DO DESENVOLVIMENTO DE SISTEMAS WEB.....	23
3 PHP & MYSQL.....	26
3.1 O PHP.....	26
3.1.1 VANTAGENS DA UTILIZAÇÃO DO PHP.....	27
3.2 O MYSQL.....	27
4 DATA MINING.....	29
4.1 FUNÇÕES DO DATA MINING.....	31
4.1.1 CLASSIFICAÇÃO.....	31
4.1.2 ESTIMAÇÃO.....	32
4.1.3 PREVISÃO OU PREDIÇÃO.....	32
4.1.4 ASSOCIAÇÃO.....	33
4.1.5 ANÁLISE DE AGRUPAMENTO.....	34
4.2 A TÉCNICA ÁRVORE DE DECISÃO.....	34
5 A IEADB.....	37
5.1 MEMBROS DA IEADB.....	38
5.2 ENTRADAS DA IEADB.....	39
5.3 SAÍDAS DA IEADB.....	40
5.4 PROBLEMA ATUAL.....	41
6 TRABALHOS CORRELATOS.....	43
6.1 ROL-BL.....	43
6.2 GESIG.....	45
7 DESENVOLVIMENTO DO TRABALHO.....	46

7.1 REQUISITOS PRINCIPAIS DO PROBLEMA	46
7.2 ESPECIFICAÇÃO	47
7.2.1 NÍVEL 1 E NÍVEL 2	48
7.2.2 NÍVEL 3.....	50
7.3 A MINERAÇÃO DE DADOS DO SISTEMA	55
7.4 IMPLEMENTAÇÃO	63
7.4.1 TÉCNICAS E FERRAMENTAS UTILIZADAS.....	64
7.4.2 OPERACIONALIDADE DA IMPLEMENTAÇÃO.....	65
7.5 RESULTADOS E DISCUSSÃO	75
8 CONCLUSÕES.....	77
8.1 EXTENSÕES	78
REFERÊNCIAS BIBLIOGRÁFICAS	79

1 INTRODUÇÃO

Percebe-se nos últimos anos no Brasil o avanço desenfreado do número de evangélicos. Segundo Gwercman (2004), “*um a cada seis brasileiros já é evangélico – e o número continua crescendo*”, sendo que há meio século os evangélicos são a religião que mais cresceram no país. Nestes últimos vinte anos, o número de evangélicos mais que triplicou, pois em 1980 o número de pessoas evangélicas era de 7,8 milhões, já em 2001 o número de pessoas evangélicas era de 26,4 milhões, sendo que enquanto a população cresceu 6,6% os evangélicos cresceram 15,6%.

No gráfico a seguir é comparado o crescimento da população em cada região do Brasil em relação ao crescimento do número de evangélicos no Brasil, do ano de 1980 até o ano de 2000 (Figura 1).



Fonte: Instituto de Geografia e Estatística (IBGE) – (Censo Demográfico 2002)

Figura 1 – Gráfico do crescimento dos evangélicos no Brasil em relação a população

Em paralelo com o crescimento do número de evangélicos, crescem também as igrejas evangélicas no país, sendo que entre as igrejas evangélicas existentes no Brasil destaca-se a Igreja Evangélica Assembléia de Deus (IEAD) que segundo o IBGE cresceu anualmente 14,8% entre o ano de 1991 até 2000, e possui atualmente cerca de 8,7 milhões de membros, número diferente do registrado pela Convenção Geral das Assembléias de Deus no Brasil (CGADB) que acredita possuir mais de 15 milhões de membros e mais de 100 mil templos espalhados pelo país. A Igreja Evangélica Assembléia de Deus de Blumenau (IEADB) possui em sua relação de membros ativos mais de 5000 pessoas cadastradas e aproximadamente 60 templos espalhados pelos bairros da cidade.

Crescendo o número de evangélicos e conseqüentemente de igrejas, verifica-se a dificuldade que estas igrejas têm para organizar a parte administrativa, controlar a relação de seus membros e controlar os processos administrativos que ocorrem em uma igreja. Existindo esta dificuldade se faz necessário um sistema que auxilie e otimize estas tarefas, segundo as necessidades específicas de cada igreja, pois cada igreja possui suas particularidades, tanto na parte eclesiástica como na parte administrativa. Portanto, para a realização deste trabalho estudou-se a forma de administração e organização da IEADB para a implementação de um sistema web utilizando mineração de dados.

Segundo Goodman (2001, p.4), com a liberação do uso comercial da internet, milhares de empresas criaram suas páginas na web: companhias aéreas, bancos, seguradoras, redes de rádio e televisão, revistas, jornais, editoras, fabricantes de software e hardware, lojas de computadores, montadoras e concessionárias de veículos, supermercados, restaurantes, livrarias, lojas de discos e muitas outras. Além de marketing institucional, essas empresas e instituições vêm na web um excelente veículo para venda, divulgação de produtos, divulgação de serviços com um grande alcance de pessoas. Então porque não, uma igreja com um sistema totalmente baseado na web, podendo descentralizar as fontes de informações e atualizando seus dados *on-line*.

Com a utilização da mineração de dados será possível descobrir significados aos dados armazenados no banco de dados, pois este significado permite a análise dos dados observando modelos, estabelecendo mecanismos e tendo novas idéias para fazer previsões e prognósticos sobre o futuro. Conforme Berry (1997, p.5) a mineração de dados é a exploração e análise, por meios automáticos ou semi-automáticos de dados para descobrir modelos e regras significativas.

1.1 OBJETIVOS

Este trabalho tem como objetivo principal a especificação e implementação de um sistema web para auxiliar e otimizar a administração da IEADB e de igrejas que tem o mesmo princípio administrativo utilizando mineração de dados para auxiliar os administradores da igreja em suas decisões.

Os objetivos específicos são:

- a) otimizar a forma de administração da igreja através do sistema;
- b) implementar o protótipo do sistema;
- c) facilitar as transações de informações entre a sede e as demais congregações;
- d) fazer validações para verificar a consistência nos dados cadastrados no banco de dados;
- e) utilizar a mineração de dados para verificar o desempenho das congregações.

1.2 ESTRUTURA DO TRABALHO

O trabalho está dividido em oito capítulos, cuja descrição segue.

O primeiro capítulo faz uma introdução, abordando de maneira geral o trabalho, sua relevância, objetivo e o contexto no qual está inserido.

O segundo capítulo apresenta uma visão geral sobre sistemas baseados na web, mostrando conceitos, vantagens, problemas e utilidades dos mesmos, pois este é o ambiente que o sistema vai utilizar.

O terceiro capítulo apresenta de forma resumida alguns conceitos da união entre PHP e MySQL.

O quarto capítulo enfatiza alguns conceitos, técnicas e aplicação da mineração de dados.

O quinto capítulo faz uma amostragem geral da atual forma de administração da IEADB, apresentando soluções para a melhoria dos processos e da administração.

O sexto capítulo apresenta alguns trabalhos correlatos existentes e utilizados atualmente.

O sétimo capítulo trata sobre as especificações e a implementação do sistema, através dos seus requisitos, diagramas de caso de uso, descrição dos casos de uso primário, diagrama entidade/relacionamento (DER) e o dicionário de dados do sistema. Neste capítulo também é feita a apresentação da interface e da utilização do sistema.

O oitavo e último capítulo completa o trabalho, apresentando as conclusões dos resultados do trabalho e as vantagens do mesmo. Também são verificadas as limitações do sistema e algumas sugestões para futuros trabalhos em relação ao apresentado.

2 SISTEMAS WEB

A *World Wide Web* e a internet colocaram as pessoas, em geral, no mundo da computação. Compra-se ações, ouve-se música, visualiza-se filmes, obtêm-se conselhos médicos, reserva-se passagens aéreas, se conhece pessoas, efetua-se transações bancárias, compra-se suprimentos – quase tudo é possível no mundo virtual da web. Segundo alguns a web e a internet são o desenvolvimento mais importante da computação, pois estas tecnologias colocaram as pessoas na idade da informação. Com esta rápida e dinâmica expansão de utilidades on-line, a internet tem aberto inúmeras novas possibilidades para a implantação e implementação de serviços computacionais. A web é, atualmente, o principal veículo para a prestação destes serviços, permitindo atingir um número cada vez maior e mais diversificado de usuários e isto torna a demanda por sistemas baseados na web cada vez maior e urgente.

Segundo Conallen (2003, p.10), o termo web vem da visão do sistema como um conjunto de nós com links de interconexão. Vendo por este lado, ele parece com a rede de uma aranha. Os links proporcionam um meio de navegar pelos recursos do sistema. A maioria deles conecta documentos de texto, mas o sistema pode ser usado também para distribuir áudio, vídeo e dados personalizados. Os links facilitam a navegação para outros documentos, pois, o usuário só precisa clicar em um link e o navegador interpreta isso como uma solicitação para carregar o documento ou recurso referenciado no seu local.

Conforme Stair (1998), sistema é um conjunto de elementos ou componentes que interagem para se atingir os objetivos. Oliveira (1992), define sistema como um conjunto de partes que, juntas formam um todo, para exercerem determinadas funções e atingirem determinados objetivos, acrescentando que os componentes de um sistema são as entradas, o processamento e saídas.

Um sistema ou aplicação web é desenvolvido para adicionar funcionalidades de negócio para alcançar o objetivo deste determinado negócio via web (CONALLEN, 2003). Nos seus termos mais simples, uma aplicação web é um sistema que permite a seus usuários executar a regra ou lógica do negócio com um navegador web. Existe uma distinção sutil entre um sistema web e um simples site web. Um sistema web é um site em que a entrada do usuário no sistema afeta o estado do negócio. Essencialmente, um sistema baseado na web usa um site web como o *front end* para aplicação de negócio. Em quase todas as aplicações web

mais simples, o usuário precisa fornecer mais do que apenas informações de solicitação de navegação. Normalmente, os usuários de um sistema web inserem uma gama variada de dados, textos simples, seleções de caixas de seleção ou, até mesmo, informações binárias ou de arquivo. Esses dados serão armazenados e de alguma forma o sistema vai ter um estado diferente em algum determinado lugar quando o usuário terminar de usar o sistema, ou seja o usuário afetou de alguma forma o estado do negócio.

A arquitetura de um sistema web é simples e direta. Ela contém os mesmos componentes principais de um site web: um servidor web, uma conexão de rede e navegadores cliente. A inclusão do servidor permite que o sistema gerencie o estado e a regra do negócio.

2.1 A WEB DINÂMICA

Atualmente a maioria das pessoas tem algum tipo de contato com a internet quer seja para envio de *e-mails*, para visualização de notas no colégio, ou para uma pesquisa de escola, e até mesmo para acessar as informações da empresa em que se trabalha. Segundo Soares (2000, p.1) a web de hoje é muita distinta da web de poucos anos atrás, quando somente se tinha acesso a páginas estáticas que sempre mostravam a mesma “cara” para todos os internautas. Bem, isto mudou e para melhor, pois hoje em dia é possível acessar sites na web que fazem de tudo um pouco, desde pesquisa em outros sites, compra de ações, carros, casas, e até mesmos empresas trabalhando com sistemas totalmente baseados na web. Enfim, grande parte das coisas que o mundo real pode ter e oferecer, provavelmente, pode ser encontrado na web.

Mas, para que tudo isto aconteça, é necessário que muitas pessoas trabalhem nos bastidores do mundo virtual, para prover os sites de vida própria. Pessoas estas que transformam páginas estáticas sem graça em verdadeiras fontes de informação ou em até mesmo sistema de negócios utilizando as ferramentas de programação para a internet. Estas ferramentas (PHP, ASP, Cold Fusion, PERL, e outros) tornam as páginas e sites estáticos em páginas e sites dinâmicos onde é possível a interação com este determinado site ou página. A partir da possibilidade de dinamizar a web transformando uma página de conteúdo estático em um conteúdo dinâmico, as empresas estão cada vez mais fazendo uma migração de seus sistemas comuns para sistemas baseados totalmente ou parcialmente na web.

2.2 CARACTERÍSTICAS DO PROCESSO DE SISTEMAS WEB

Segundo Pressman (2002, p.754) as seguintes características guiam o processo de sistemas baseados na web:

- a) imediatismo: aplicações baseadas na web têm um imediatismo que não é encontrado em nenhum outro tipo de software. Isto é, o prazo de colocação no mercado e disponibilização de novas informações de um site pode ser uma questão de semanas. Os desenvolvedores precisam usar métodos para planejamento, projeto, implementação e testes;
- b) segurança: como as aplicações web estão disponíveis através de acesso à rede, é difícil limitar a população de usuários finais que podem ter acesso à aplicação. A fim de proteger o conteúdo reservado e fornecer modos seguros de transmissão de dados, boas medidas de segurança precisam ser implementadas na infra-estrutura da aplicação propriamente dita;
- c) estética: uma inegável parte da atração de uma aplicação ou sistema web é o seu aspecto. Quando uma aplicação é projetada para o mercado, para vender produtos e idéias, ou para que os usuários se sintam bem e fiquem à vontade no uso da aplicação a estética pode ter tanto a ver com o sucesso quanto o projeto técnico.

2.3 PRINCIPAIS VANTAGENS DE SISTEMAS WEB

A seguir estão citadas algumas vantagens da utilização dos sistemas que são baseados na plataforma web:

- a) sistemas baseados na web podem ser executados a partir de qualquer navegador da internet;
- b) sistemas baseados na web podem ser acessados de qualquer lugar do mundo, para isto basta apenas o usuário possuir um computador com conexão à internet e um navegador, pois o sistema web fica acessível em qualquer computador seja na rede local ou na web facilitando aos usuários o acesso ao aplicativo onde quer que estejam. Imagine um cenário onde o gerente de um determinado departamento esteja de férias e este departamento necessite solicitar ao almoxarifado uma mesa e algumas cadeiras novas e que apenas o gerente tem autorização para proceder com tal pedido. Basta este gerente conectar na internet através de um *cyber* café ou escritório virtual e efetuar o pedido entrando com sua senha. Ou seja, ele não

- precisa passar sua senha para outro funcionário, não precisa estar fisicamente na empresa e não precisa delegar este tipo de poder temporariamente a ninguém;
- c) interface HTML reconhecida por uma grande gama de usuários já acostumados com o funcionamento dos navegadores;
 - d) atualização dos dados e informações em tempo real para todos os usuários do sistema;
 - e) desenvolvimento, manutenção e atualização centralizada da aplicação. Não é necessário sair instalando o sistema em diversos equipamentos diferentes. Basta colocá-lo no servidor para que os usuários obtenham acesso, gerando minimização dos custos, pois em qualquer situação, seja de atualização e ou alteração do sistema basta fazer apenas no servidor e a partir de então todos os usuários do sistema desfrutarão das mudanças efetuadas;
 - f) redução dos custos de comunicação, sendo que se existem escritórios dispersos e o sistema de informação se baseia na internet, o custo em conversações pode ser substancialmente reduzido;
 - g) exportação de dados entre usuários remotos usando o protocolo HTTP é mais simples e mais fácil do que usar outro protocolo;
 - h) não é exigido muita memória e nem poderosos processadores para a execução do sistema nos terminais pois o sistema é todo executado no servidor;
 - i) escalabilidade no processamento. Se houver necessidade de aumentar o poder de processamento, basta fazer isto no servidor;
 - j) pode proporcionar potencialmente, uma melhor produção para a empresa pois o funcionário pode trabalhar em qualquer local onde o acesso a internet é disponibilizado.

2.4 PRINCIPAIS DESVANTAGENS DE SISTEMAS WEB

A seguir estão citadas algumas desvantagens da utilização dos sistemas que são baseados na plataforma web:

- a) não há uma padronização entre os diversos navegadores e o sistema poderia ser exibido de uma maneira diferente dependendo do navegador e da versão deste navegador também;
- b) a entrada de uma grande massa de dados é prejudicada na interface HTML pois não existe uma maneira padrão de criar máscaras de entrada de dados;

- c) tempo de processamento da execução das tarefas depende da velocidade da conexão, entre cliente e servidor;
- d) os sistemas baseados na web dependem dos recursos do navegador usado para visualizar a aplicação. Como eles possuem recursos diferentes, existem dificuldades para prever como a aplicação vai se comportar;
- e) desenvolver páginas dinâmicas e formulários com interface HTML para entrada de dados é muito mais trabalhoso e complicado que em aplicações comuns;
- f) a manipulação das variáveis é um trabalho muito mais complicado, tendo em vista a possibilidade que o usuário tem de abrir e fechar janelas e “navegar” para onde bem entender;
- g) desenvolvimento mais complicado pois envolvem três camadas onde é necessário o servidor de banco de dados, servidor de aplicativos (regras de negócios) que será o servidor de internet e o *front end* com diversas validações no próprio navegador;
- h) a interface HTML não é rica em controles gráficos e peca no quesito posicionamento. O visual da aplicação pode não ficar tão elegante em relação ao momento do desenvolvimento;
- i) difícil gerenciamento do estado do cliente no servidor. A natureza sem conexão das comunicações do cliente e do servidor não proporciona um modo fácil do servidor controlar a solicitação de cada cliente e associá-la à solicitação anterior, visto que cada e toda solicitação de página web estabelece e, em seguida, interrompe um conjunto completamente novo de conexões;

2.4.1 SEGURANÇA

Uma das principais desvantagens e preocupações de um sistema baseado na web é o quesito segurança. Mesmo que o sistema seja para uma *intranet*, protegida por um *firewall* da empresa, a segurança continuará sendo uma preocupação. Segurança é o termo que usa-se para descrever a proteção dos dados e sistema. Um sistema seguro é uma aplicação que funciona adequadamente e que faz apenas o que se propõe a fazer, sem comprometer a integridade dos dados para aqueles que não estão autorizados a obter determinadas informações (CONALLEN, 2003).

Pessoas sem escrúpulos mesmo com acesso limitado ao sistema, podem aproveitar qualquer falha do sistema para obter acesso a informações potencialmente valiosas, como

perfis de usuário, número de cartão de crédito, ou simplesmente derrubar o sistema para testar sua perícia e orgulho pessoal. A ameaça é muito real e com os sistemas baseados na web assumindo papéis cada vez mais importantes, a necessidade de entender e administrar os riscos de segurança se torna ainda mais crítica.

Para entender as áreas de risco de um sistema baseado na web, é necessário entender primeiramente onde os sistemas são vulneráveis. A arquitetura de sistemas web básica, sendo uma variante de uma arquitetura cliente/servidor, tem três elementos arquitetônicos principais: o cliente, o servidor e a rede. Cada um deles é vulnerável a ataques.

- a) os clientes correm o risco de ataques de softwares que danificam o sistema do cliente ou comprometem os recursos do cliente particular como informações pessoais e arquivos;
- b) os servidores correm o risco de acesso não-autorizado, o qual pode resultar na captura de informações confidenciais, na execução de programas prejudiciais no servidor ou, ainda, desativar temporariamente as funções do servidor;
- c) as redes podem ser monitoradas e as comunicações de dados entre o cliente e o servidor podem ser interceptadas.

É tarefa do arquiteto e dos designers chefes entenderem e administrarem esses riscos com determinadas estratégias e tecnologias para evitarem qualquer tipo de problema em relação à segurança. Algumas recomendações de estratégias de segurança são citadas por Conallen (2003, p.96) na modelagem de um sistema seguro:

- a) controle de acesso: limitar a funcionalidade do sistema a usuários específicos;
- b) autenticação: identificar alguém que tenha direitos de acesso ao sistema. Relacionada à autenticação está a identificação, que é a tentativa de determinar uma identidade específica, na maioria dos sistemas baseados na web esta atividade é efetuada com a utilização de sessões;
- c) auditoria: registrar as atividades do sistema, especialmente a atividade dos usuários do sistema;
- d) detecção de intruso: detectar usuário não-autorizado do sistema;
- e) criptografia: converter dados em um formato que não possa ser facilmente entendido por aqueles que não têm autorização para visualizá-los.
- f) manutenção da segurança do sistema: gerenciar medidas defensivas do sistema e outras funções que estão relacionadas à segurança.

2.4.2 FALTA DE PADRÃO NA WEB

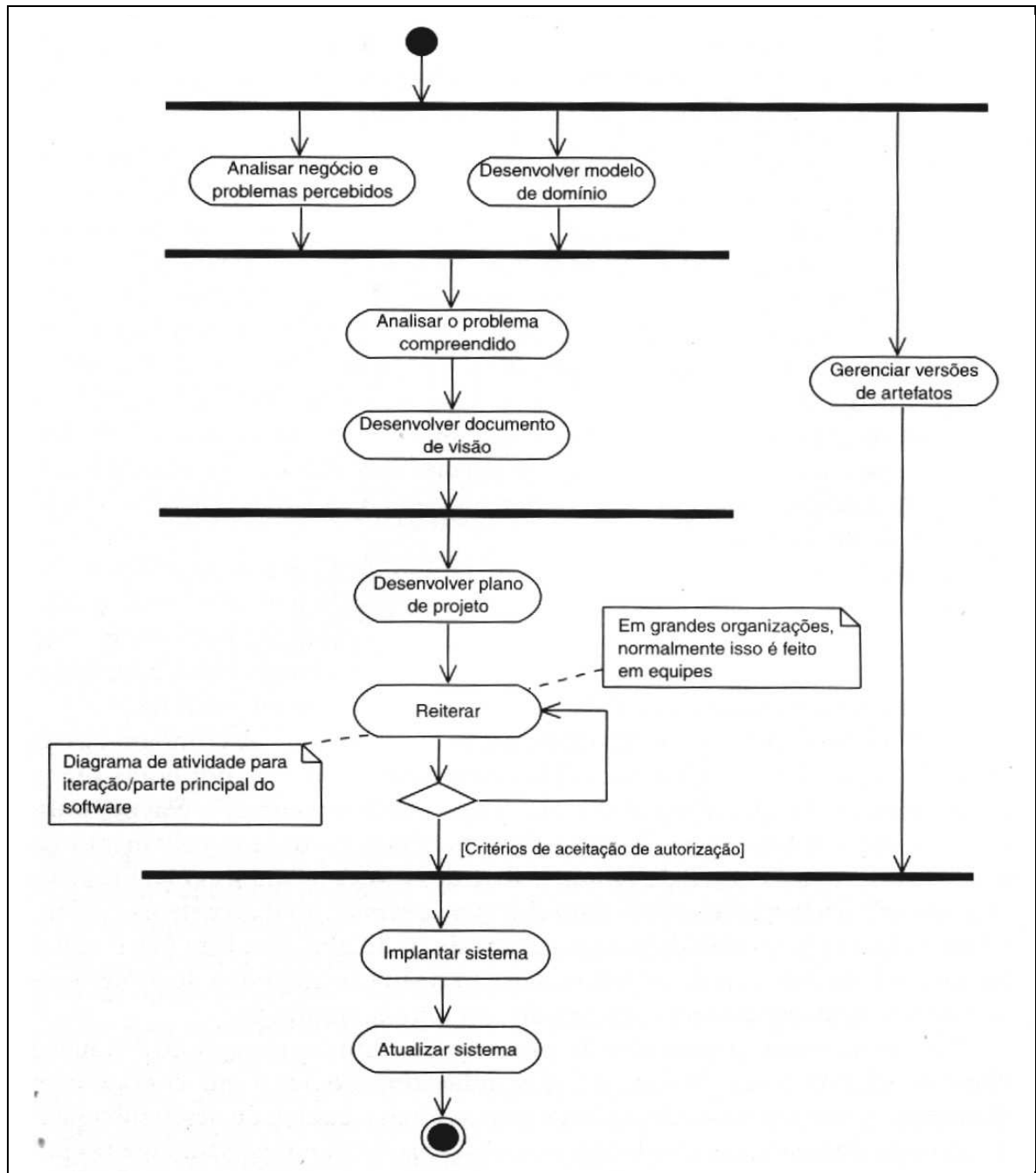
A falta de padrões para o desenvolvimento de sites e sistemas na web ainda é um assunto muito discutido entre os desenvolvedores web. A divergência começa desde a escolha do navegador a ser utilizado, até o formato, estilo e aparência que a página web vai possuir. Para amenizar este problema, em 1994 o World Wide Web Consortium (W3C) foi criado. Essa organização tem como objetivo guiar a criação de padrões para a utilização em documentos Web. Obviamente, em 1994 a situação já era caótica e mesmo apesar do esforço desenvolvido pela organização nos últimos anos a situação atual ainda está longe de ser perfeita. Ainda assim, se o W3C não existisse, hoje seria impossível criar documentos que se comportassem de maneira mais ou menos similar nos vários navegadores e sistemas atuais (DEITEL, 2003).

A falta de padrão na web se torna “um terror” para os desenvolvedores web, pois estes vêem o risco de seus scripts e rotinas não serem executados corretamente após o desenvolvimento, o que causa transtorno e muita frustração ao desenvolvedor. Por isso duas das recomendações da W3C é escrever os códigos com suporte a mais de um navegador e seguir a instrução “*keep it simple*” ou seja mantenha isto simples.

O W3C dedica-se a desenvolver tecnologias interoperáveis e não-proprietárias para web. Um dos objetivos principais da W3C é tornar a web universalmente acessível. As tecnologias de web padronizadas pelo W3C são chamadas de recomendações. As recomendações da W3C incluem HTML, CSS (*Cascading Style Sheets*), HXML e XML. Uma recomendação não é um produto real de software, mas um documento que especifica o papel, a sintaxe e as regras de uma tecnologia.

2.5 ATIVIDADES DO DESENVOLVIMENTO DE SISTEMAS WEB

Segundo Conallen (2002, p.110) estas são as principais atividades a serem tomadas no desenvolvimento de um sistema web (Figura 2):



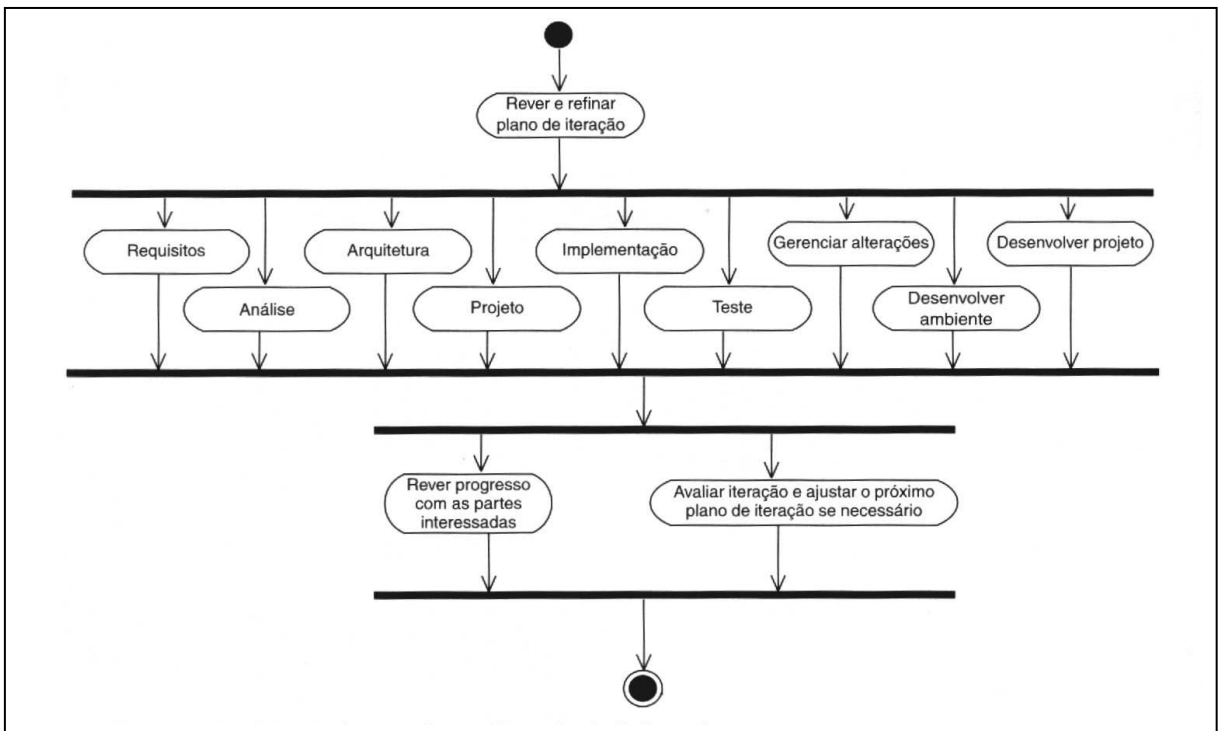
Fonte: Conallen (2003)

Figura 2 – Diagrama de atividade do caso de uso do desenvolvimento do software

- a) gerenciar versões de artefatos: essa atividade é feita simultaneamente com todas as outras atividades que envolvem a evolução e a criação dos artefatos do processo.

Essencialmente, é a atividade do processo de gerenciamento de alterações e o uso de um sistema de controle de versão;

- b) analisar negócio e problemas percebidos: adota uma análise objetiva do estado do negócio. Procurando não ser influenciado pelos problemas percebidos, quando apresentados pelas partes interessadas, pois estes podem ser apenas superficiais e ocultam os problemas reais, que estão muito mais profundamente arraigados. Somente com uma compreensão do negócio é possível examinar os problemas reais de um modo verdadeiramente imparcial;
- c) desenvolver modelo de domínio: utilizar-se das ferramentas de desenvolvimento de software, como a UML, a reunião de requisitos e a aplicações de gerenciamento de documentos, para construir um modelo de negócio. O artefato mais importante do modelo de domínio, entretanto, é o glossário, que define os termos e conceitos principais do contexto no qual o sistema está trabalhando;
- d) analisar o problema compreendido: com um sólido entendimento do estado do negócio e do domínio em que ele opera, é hora de focar o problema real ou pelo menos o problema em que as partes interessadas estão dispostas a permitir o trabalho do desenvolvedor;
- e) desenvolver documento de visão: o documento de visão, a partir do qual todos os outros derivam, expressa o escopo e a finalidade de todo o projeto de software. Normalmente, o documento de visão se concentra em fornecer os recursos principais e critérios de aceitação para o sistema em desenvolvimento;
- f) desenvolver plano de projeto: o plano de projeto esboça as atividades do esforço de desenvolvimento de todo o software, definido os eventos principais e citando documentos padrões apropriados, incluindo o plano de gerenciamento de alteração e de configuração;
- g) reiterar/iteração de software: aqui está a parte principal do software, onde são descritas as atividades do processo de desenvolvimento que envolve a iteração de todos os artefatos que compõem os pontos de partida do documento de visão para os artefatos que compõem o produto final. Na Figura 3 é ilustrado o diagrama de atividade do caso de uso “iteração de software”.



Fonte: Connalen (2003)

Figura 3 – Diagrama de atividades do caso de uso iteração de software

- h) implantar sistema: a produção e instalação do sistema pode envolver um lançamento do produto planejado em fases ou significar simplesmente ativá-lo. Se envolvidos em um lançamento planejado em fases, as atividades e os fluxos de trabalho devem ser concluídos durante a fase de construção;
- i) atualizar sistema: a atualização do sistema é essencialmente uma mini-versão do processo que o desenvolveu.

3 PHP & MYSQL

A junção da linguagem PHP e do banco de dados MySQL vem recebendo a cada dia um grande número de adeptos, pois oferece diversas vantagens aos desenvolvedores como também às empresas e organizações que utilizam sites e sistemas desenvolvidos a partir destas tecnologias (SOARES, 2001).

3.1 O PHP

PHP é uma linguagem de código aberto que executa no lado servidor permitindo a criação de sites web dinâmicos, possibilitando uma interação com o usuário através de formulários, parâmetros da URL e links (CONVERSE, 2001).

A linguagem PHP foi concebida em 1994 por Rasmus Lerdorf. As primeiras versões não foram disponibilizadas, tendo sido utilizadas apenas na sua *home page* para que ele pudesse ter informações sobre as visitas que estavam sendo feitas. A primeira versão utilizada por outras pessoas foi disponibilizada em 1995, e ficou conhecida como *Personal Home Page Tools* (ferramentas para página pessoal). Era composta por um sistema bastante simples que interpretava algumas macros e alguns utilitários que rodavam “por trás” das páginas pessoais como um livro de visitas e um contador.

Em meados de 1995 o interpretador foi reescrito e ganhou o nome de PHP/FI. O “FI” veio de um outro pacote escrito por Rasmus que interpretava dados de formulários HTML (*Form Interpreter*). Ele combinou os scripts do pacote *Personal Home Page Tools* com o FI e adicionou suporte a mSQL (mini SQL), nascendo assim o PHP/FI, que cresceu bastante, e as pessoas passaram a contribuir com o projeto.

Estima-se que em 1996 PHP/FI estava sendo usado por cerca de 15.000 sites pelo mundo, e em meados de 1997 esse número subiu para mais de 50.000. Nessa época houve uma mudança no desenvolvimento do PHP. Ele deixou de ser um projeto de Rasmus com contribuições de outras pessoas para ter uma equipe de desenvolvimento mais organizada. O interpretador foi reescrito por Zeev Suraski e Andi Gutmans, e esse novo interpretador foi a base para a versão 3.

A diferença de PHP com relação a linguagens semelhantes a Javascript é que o código PHP é executado no servidor, sendo enviado para o cliente apenas HTML puro. Desta

maneira é possível interagir com bancos de dados e aplicações existentes no servidor, com a vantagem de não expor o código fonte para o cliente. Isso pode ser útil quando o programa está lidando com senhas ou qualquer tipo de informação confidencial.

3.1.1 VANTAGENS DA UTILIZAÇÃO DO PHP

Algumas vantagens da utilização da linguagem PHP de acordo com Segundo (2004):

- a) código aberto (*open source*): significa que está disponível para o uso público geral sem custo de licença de uso;
- b) amplamente utilizada (*widely use*): significa que é popular e muitos desenvolvedores web estão utilizando;
- c) código fonte de fácil compreensão;
- d) rápida apesar de interpretada;
- e) pode-se optar pelo controle centralizado ou distribuído;
- f) acesso à diversos banco de dados, como MySQL, PostgreSQL, Oracle, dBase, InterBase, Informix, Sybase, Ingres, IBM DB2 e outros;
- g) suporte a orientação à objetos, embora ainda não suporta todos os recursos da orientação à objetos;
- h) possui diversos recursos como XML, FTP, PDF, sessões, expressões regulares, biblioteca matemática, criptografia e outros.

3.2 O MYSQL

O programa MySQL é um servidor robusto de bancos de dados SQL (*Structured Query Language*), multi-tarefa e multi-usuário. Escrito em C e em C++, originou-se no início dos anos 80 e tem sido usado em projetos desde 1996. Foi exaustivamente testado por milhões de usuários da comunidade *open source* com diversos compiladores em diferentes plataformas. Todos os erros reportados pelos usuários são corrigido em um novo lançamento – a solução de erros mais sérios é implementada em novas versões (LIMA, 2003).

O servidor MySQL pode ser usado em sistemas de produção com alta carga e missão crítica bem como pode ser embutido em programa de uso em massa. O MySQL é um dos mais utilizados sistemas de gerenciamento de banco de dados relacional *open source* e sem dúvida o mais utilizado com a linguagem PHP para aplicações baseadas na web.

Apesar de possuir algumas restrições em relação a banco de dados *open sources* como PostgreSQL e como o Oracle que não é *open source*, seguem algumas características do que o MySQL faz de melhor segundo Suehring (2002, p.21):

- a) multi-plataforma: funcionamento em diversas plataformas, como Win32, Linux, FreeBSD, Unix, e outros;
- b) aplicações web: as aplicações web em geral apresentam muitas leituras e poucas gravações. O MySQL é rápido e pode atender às demandas de velocidade na internet. Nas experiências realizadas o MySQL tem provado repetidas vezes que supera outros produtos de Sistemas Gerenciadores de Banco de Dados Relacional (SGBDR) em aplicações web;
- c) aplicações de nível corporativa: o MySQL oferece suporte diretamente pela empresa pai, MySQL AB. O conjunto de recursos do MySQL inclui quase tudo que uma aplicação de nível corporativo precisaria;
- d) suporte a código-fonte aberto: o MySQL também responde bem às solicitações de recursos. O MySQL tem seu código-fonte aberto; todo mundo é convidado a fazer *download* dele e estender o código para atender suas próprias necessidades;
- e) sobrecarga baixo: o MySQL executa confortavelmente muitas aplicações em um computador pessoal;
- f) tamanho grande de tabela disponível: as tabelas do MySQL podem crescer bastante, embora às vezes encontrem limitações de tamanho de arquivo do sistema operacional da hospedagem. Algumas arquiteturas, porém, podem acomodar até oito terabytes por tabela utilizando MySQL;
- g) estabilidade: todo software está em desenvolvimento. Porém recursos como tipo de tabela padrão, a funcionalidade básica SQL, software do cliente APIs Perl e PHP, são consideradas totalmente estáveis;
- h) baixo TCO (*Total Cost of Ownership*): oferece economia de recursos em geral, gerenciando mais por menos;
- i) gratuito: assim como o PHP a grande vantagem do MySQL é ser *open source*, isto significa estar disponível para o uso público geral com custo zero;
- j) padronização: o MySQL segue quase inteiramente o padrão SQL-92.

4 DATA MINING

Nas duas últimas décadas aconteceu um aumento significativo na quantidade de informações ou dados que são armazenadas em formato eletrônico. Foi calculado que a quantidade de informação no mundo dobra a cada vinte meses e o tamanho e o número de bancos de dados está aumentando ainda mais rapidamente. O valor destes dados armazenados está tipicamente ligado a capacidade de extrair informações de mais alto nível que se encontra subjacente a estes dados, ou seja, informação útil que sirva para dar suporte a decisões, e para exploração e melhor entendimento do fenômeno gerador dos dados. Podem existir padrões ou tendências úteis interessantes que, se descobertos, podem ser utilizados, por exemplo, para otimizar um processo de negócio em uma empresa, ajudar no entendimento dos resultados de um experimento científico, entre outros (ARBEX, 2003).

Conforme Fayyad et al. (1996, p.27), *data mining* é o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e compreensíveis.

A expressão *data mining* (DM), mineração de dados ou garimpagem de dados refere-se ao processo de extrair e explorar informações potencialmente úteis a partir de dados que estão armazenados nos bancos de dados. A tecnologia utilizada na mineração de dados procura extrair padrões e relacionamentos que podem ser fundamentais para o negócio. O DM trabalha com um conjunto de técnicas avançadas e princípios de inteligência artificial para identificar os padrões e associações que os dados refletem, com isso oferecendo conclusões que podem trazer valiosas vantagens de mercado para empresa ou para o negócio em questão (HARRISON, 1998).

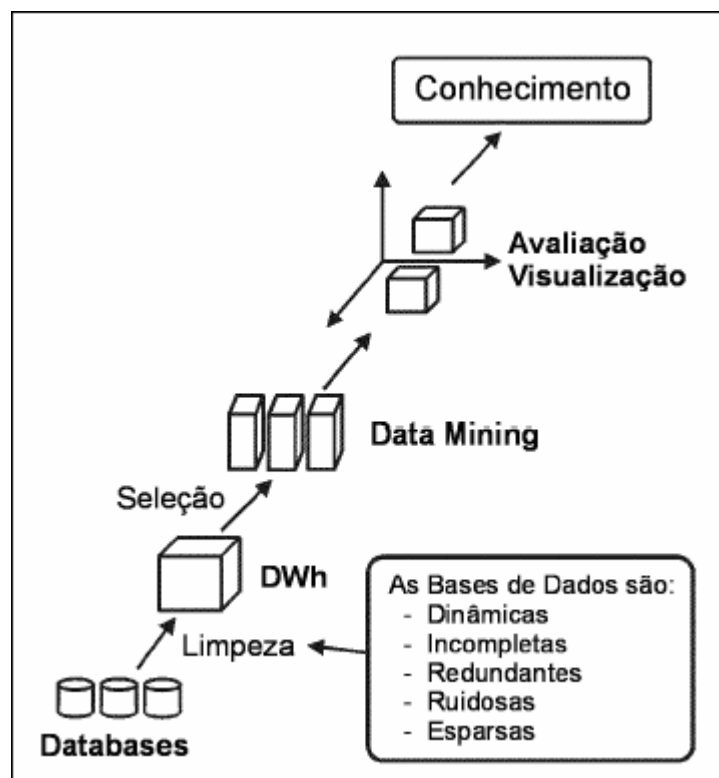
O processo de descobrimento realizado pela mineração de dados pode ser utilizado a partir dos sistemas transacionais, mas é mais eficiente quando utilizado a partir de um *data warehouse* onde os dados já estão mais consistentes e íntegros, e habilitam descobertas mais abrangentes e precisas. Este processo de DM permite investigar esses dados à procura de padrões que tenham algum valor ou significado para a empresa ou negócio (LOH, 1997).

Segundo Mena (1999, p.4), DM é o processo de descobrir informações relevantes, como padrões, associações, mudanças, anomalias e estruturas, em dados armazenados em banco de dados, depósitos de dados ou outros depósitos de informação. Mecanismo de transformação de dados de baixo nível em informação de alto nível, ajudando no processo de

tomada de decisões organizacionais através do uso de técnicas automáticas ou semi-automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações, que devido a esse volume não seriam descobertas a olho nu.

A mineração de dados é o processo mais utilizado para extração de conhecimento através de bancos de dados, este processo é chamado de KDD (*Knowledge Discovery in Databases*), tanto no meio comercial quanto no meio científico. O termo KDD refere-se ao processo global de descobrimento de conhecimento útil em bases de dados. O processo de descoberta do conhecimento (Figura 4) divide-se em:

- a) preparação: os dados são selecionados, purificados e pré-processados;
- b) mineração: os dados são processados e informações importantes são identificadas;
- c) análise de dados: verifica-se as informações adicionais que foram descobertas e a importância dos fatos gerados, para depois criar um método de visualização como um gráfico ou uma tabela.



Fonte: (NAVEGA, 2002)

Figura 4 – Processo de descoberta do conhecimento através do banco de dados

A extração de conhecimentos em bases de dados é um processo complexo e, ainda hoje, muito dependente da experiência e do trabalho do analista (formulação do problema,

preparação dos dados, análises dos resultados, avaliações), por isso é indispensável a presença do mesmo. É atribuído às máquinas a responsabilidade de manipular conjuntos de dados, procurando sempre de maneira eficaz, padrões e resultados que satisfazem os problemas apresentados. Neste processo utiliza-se um conjunto de técnicas estatísticas e também inteligência artificial.

4.1 FUNÇÕES DO DATA MINING

O DM pode desempenhar uma série de tarefas dependendo das circunstâncias em que é aplicada. Cada classe de aplicação no DM tem como base um conjunto de algoritmos que serão usados na extração de relações relevantes dentro de uma massa de dados. Segundo Westphal (1998, p.173) as funções mais utilizadas são:

- a) classificação;
- b) estimação;
- c) previsão ou predição;
- d) associação;
- e) análise de agrupamento.

4.1.1 CLASSIFICAÇÃO

Classificação é a tarefa mais utilizada em DM, pois serve para determinar com que grupos de entidades já classificadas (exemplos) anteriormente o dado apresenta semelhança, ou seja, classifica ou associa um registro em classes predefinidas, determinando com que grupo de entidades já classificadas determinado dado apresenta mais semelhança.

Como exemplo, pode ser utilizado para classificar os clientes de uma empresa, segundo o tipo de compras (boa, média e ruim) e os atributos dos produtos (preço, marca, posição na loja, tipo e categoria). Se a idade do cliente é inferior a 18 anos e o sexo do cliente é masculino então o produto comprado é um videogame.

Algumas técnicas mais comuns para implementar a classificação são redes neurais, árvores de decisão e estatística (análise combinatória e regressão lógica).

4.1.2 ESTIMAÇÃO

Enquanto a classificação lida com valores discretos, a estimativa trabalha com valores numéricos contínuos. Dadas informações de entrada, a tarefa de estimar é buscar um valor numérico para uma variável contínua, como renda, altura, ou balanço de cartão de crédito.

Estimar uma grandeza é avaliá-la tendo como base casos semelhantes nos quais essa esteja presente. A estimação prevê um valor desconhecido com base em outros valores de situações idênticas, porém não exatamente iguais. Ao invés de um classificador binário determinar um risco “positivo” ou “negativo”, a técnica gera valores de “escore”, dentro de uma determinada margem. A abordagem de estimativa tem a grande vantagem de que os registros individuais podem ser agora ordenados por classificação, e as redes neurais são adequadas a esta tarefa (COMPOLT, 1999).

Exemplos de estimação incluem:

- a) estimar a demanda do consumo de um novo produto em função da despesa feita;
- b) estimar a renda total de uma família;
- c) estimar o valor em tempo de vida de um cliente;
- d) estimar o número de filhos numa família.

Algumas técnicas mais comuns para implementar a estimação são redes neurais, algoritmos genéticos, estatística (intervalos de confiança, intervalos de predição (regressão)).

4.1.3 PREVISÃO OU PREDIÇÃO

A previsão é o mesmo que classificação ou estimativa, exceto pelo fato de que os registros são classificados de acordo com alguma atitude futura prevista. Em um trabalho de previsão, o único modo para confirmar a precisão da classificação é esperar para ver, pois a previsão determina o futuro de uma grandeza com base em comportamentos passados.

A previsão é uma tarefa variante do problema de agrupamento por afinidades, onde as regras encontradas entre as relações podem ser usadas para identificar seqüências interessantes, que serão utilizadas para predizer acontecimentos subseqüentes. Nesse caso, apenas a coexistência de itens dentro de cada transação é importante, mas também a ordem em que aparecem, e o intervalo entre elas. Seqüências podem ser úteis para identificar padrões temporais. Por exemplo, entre compras em uma loja, ou utilização de cartões de crédito, ou ainda tratamentos médicos (COMPOLT, 1999)

Exemplos de tarefas de previsão:

- a) prever os clientes que sairão nos próximos seis meses;
- b) prever qual será a população nacional daqui a dez anos;
- c) prever a quantia de dinheiro que um cliente utilizará caso seja oferecido a ele um certo limite de cartão de crédito.

Algumas técnicas mais comuns para implementar a previsão são redes neurais, árvores de decisão, estatística (regressão múltipla, regressão logística binária).

4.1.4 ASSOCIAÇÃO

A associação procura estabelecer regras que interligam um conceito a outro. O objetivo de minerar regras de associação é encontrar todos os conjuntos de itens que freqüentemente ocorrem de forma conjunta na base de dados e formar regras a partir destes conjuntos.

Quanto ao tipo de descoberta, as regras de associação são consideradas um método não supervisionado, e obtém o segundo lugar em percentual de utilização em aplicações, somente atrás de classificação.

As regras de associação são representadas da seguinte forma: $X \Rightarrow Y$ (lê-se X implica em Y), onde X é o antecessor e Y o conseqüente e X e Y são dois conjuntos de itens distintos na base de dados.

Cada regra da forma $X \Rightarrow Y$ possui dois atributos que determinam sua validade no conjunto de dados e também limitam a quantidade de regras extraídas: o suporte e a confiança. Estes possibilitam o descarte das regras julgadas de pouco interesse, já que são menos freqüentes e confiáveis.

A utilidade deste procedimento é muito grande, conforme pode ser visto nos exemplos abaixo:

- a) achar todas as regras que tenham "coca-cola dietética" como conseqüentes. Isto irá auxiliar no planejamento de lojas para vender melhor este produto (privilegiam-se os antecedentes dessas regras);
- b) achar todas as regras que tenham "iogurte" no antecedente. Isto irá auxiliar na determinação do impacto nas receitas, caso este produto seja retirado das prateleiras;

- c) achar todas as regras com "salsicha" no antecedente e "mostarda" no conseqüente. Isto irá auxiliar na obtenção de melhores regras para determinar quais os itens que devem ser vendidos em conjunto com salsichas para aumentar as vendas de mostarda.

A técnica mais comum para implementar a associação são as regras de associação.

4.1.5 ANÁLISE DE AGRUPAMENTO

A análise de agrupamento ou clusterização é a tarefa de segmentação de uma população heterogênea em um número mais homogêneo de subgrupos ou clusters, ou seja, agrupa os itens que possuem características em comum. Os registros são agrupados de acordo com a semelhança e depende do usuário determinar qual o significado de cada segmento, caso exista algum. O que distingue a clusterização da classificação é que não existem classes pré-determinadas. A clusterização é seguidamente feita como prelúdio para outras técnicas de *data mining* (MENA, 1999).

Agrupar é simplesmente classificar uma massa de dados em classes desconhecidas a priori em número e forma. Técnica preliminar normalmente usada quando pouco se sabe sobre os dados (descoberta não supervisionada de relações) é muito usada para segmentação de mercado.

Exemplos da utilização de análise de grupo:

- a) clientes divididos em grupos com estratégias de marketing distintas;
- b) agrupar clientes em conjuntos homogêneos de acordo com seus hábitos de consumo;
- c) agrupar pessoas em conjuntos homogêneos de acordo com seus hábitos de compras.

Algumas técnicas mais comuns para implementar a clusterização são redes neurais, estatística (análise de conglomerados (*cluster analysis*)) e algoritmos genéticos.

4.2 A TÉCNICA ÁRVORE DE DECISÃO

Neste trabalho utiliza-se a função de classificação da mineração de dados através da técnica de árvore de decisão pois segundo Berry (1997, p.243) as árvores de decisão são uma poderosa e popular ferramenta para classificar dados. As árvores de decisão são usadas para

descobrir regras e relacionamentos. Em uma árvore de decisão existem dois tipos de atributos, o atributo alvo (decisivo) que é aquele que contém o resultado o qual se deseja obter e os não decisivos que contém os valores que conduzem a uma decisão.

Aprendizagem com árvores de decisão é um dos métodos mais utilizados para inferência indutiva. Neste método, a função de aprendizagem é representada por uma árvore que classifica instâncias desde a raiz da árvore até alguma de suas folhas que provê a classificação para a instância (MITCHELL, 1997).

Cada nó da árvore especifica um teste para algum atributo da instância, e cada galho a partir de um nó corresponde a um possível valor deste atributo. Uma determinada instância começa a ser testada a partir da raiz da árvore, descendo pelos galhos da árvore, sendo testada a cada nó até chegar à raiz. Através de uma fórmula matemática, denominada entropia (Figura 5), são realizados cálculos sobre os atributos não decisivos, onde é escolhido um nó inicial também chamado de nodo raiz. A partir deste nó é realizado uma série de novos cálculos com o objetivo de decidir a estrutura de formação da árvore a ser gerada. Este processo é repetido até que todos os atributos estejam perfeitamente classificados ou já se tenha processado todos os atributos.

Os principais algoritmos que implementam as árvores de decisão são ID3, C4.5 e PERT, sendo que os algoritmos C4.5 e PERT são um aperfeiçoamento do algoritmo ID3 com alguns conceitos de podagem de árvore (técnicas utilizadas para cortar nós da árvore que não são potencialmente úteis) e preocupação com a performance em relação ao tempo de processamento (COMPOLT, 1999).

O ID3 é um algoritmo introduzido por Quinlan (1993) para induzir as árvores de decisão a partir dos dados. O ID3 constrói a árvore de cima para baixo, colocando o melhor atributo da instância na raiz. Os galhos que partem da raiz são alocados para os possíveis valores do atributo, e os exemplos de treinamento são classificados pelos respectivos galhos da raiz. Os demais nós a partir da raiz também são alocados para outros atributos, de acordo com a sua importância, de cima para baixo na árvore.

Para determinar que atributo deve ser representado na raiz da árvore de decisão e nos demais nós, cada atributo da instância deve ser calculado (Figura 5) usando testes estatísticos que determina o quanto ele, sozinho, influencia na classificação dos exemplos de treinamento

(ganho de informação). O melhor atributo é selecionado e usado na raiz da árvore. Este processo é repetido usando exemplos de treinamento associados com cada nó descendente para selecionar o melhor atributo a ser testado em cada ponto da árvore.

Entropia(S) = $\sum -p(I) \log_2 p(I)$ onde

\log_2 é o logaritmo de um número na base 2

$p(I)$ é a quantidade de ocorrências para cada valor possível de uma classe dividido pela quantidade total da classe

Ganho (S,A) = Entropia(S) - $\sum ((|S_v|) / |S|) * \text{Entropia}(S_v)$ onde

\sum é cada valor possível de todos os valores do atributo A

S_v é a quantidade de ocorrências de cada atributo definido por A

$|S_v|$ é o número total de elementos definido por S_v

$|S|$ é o número total de elementos da coleção

Fonte: adaptado de Quinlan (1993)

Figura 5 – Fórmulas para calcular entropia e ganho

5 A IEADB

A IEADB foi fundada em Blumenau no ano de 1929 pelo missionário Wilhelm Dreffurt, vindo da Alemanha que naquela época teve muitas dificuldades para dar início a igreja (SANTOS, 1996). Mas o tempo foi passando e os membros foram aumentando sendo que atualmente a IEADB possui mais de 5.000 membros ativos e cerca de 60 congregações espalhadas pela cidade de Blumenau.

Cada igreja possui peculiaridades na forma de administração e de organização. Por este motivo tomou-se como exemplo a forma de administração da IEADB. A IEADB possui quinze setores de trabalho divididos de acordo com a região da cidade. Os setores são constituídos de duas ou mais congregações como é visualizado no Quadro 1:

Setor 01	Sede	Setor 08	Betânia
	Boa Vista		Margem Esquerda
	Velha João Pessoa		Monte Sião
	Vila Nova		Pedro Krauss
			Vorstadt
Setor 02	Garcia	Setor 09	Cohab
	Glória		Fidélis
	Itapuí		Morro da Laguna
	Zendron		Pérola do Vale
			Três Coqueiros
Setor 03	Badenfurt	Setor 10	Betesda
	Cidade Jardim II		Pôr do Sol
	Itoupavazinha		Via Moinho
	Morell		Vila Itoupava
	Salto do Norte		
	Balsa		
Setor 04	Fortaleza	Setor 11	Velha Grande
	Rua das Missões		Dona Edite
	Moriá		Hermam Kratz
	Nova Esperança		Muth
	Vale		
Setor 05	Escola Agrícola	Setor 12	Progresso
	Coripós		Filadélfia
	Ribeirão Branco		Gustavo Mayer
	Salto Weissbach		Jordão
Setor 06	Velha Central	Setor 13	Jerusalém
	Betel		Água Verde
	Hermam Barthel		Eça de Queiróz
	Ristow		
Setor 07	Itoupava Central	Setor 14	Araranguá
	América do Sol		Morro da Garuva
	Botuverá		Rua Caçapava
	Ebenezer	Setor 15	Itoupava Norte
	Frederic Jensen		Monte Hermom
	Milano		

Quadro 1 – Setores e congregações da IEADB

Cada setor possui um responsável que é o mesmo dirigente da congregação sede do setor, ou seja, o dirigente da congregação Garcia é o responsável pelo setor 02, o dirigente da congregação Itoupava Norte é responsável pelo setor 15 e assim por diante. Os responsáveis pelos setores de trabalho são pastores auxiliares que auxiliam o pastor presidente.

Cada congregação possui um dirigente que é responsável por todos os assuntos pertinentes a sua congregação.

A IEADB é dividida também em vários departamentos descritos a seguir:

- a) igreja: é considerado o departamento soberano. A partir deste departamento se derivam todos os outros;
- b) adolescentes: responsável pelos adolescentes da igreja;
- c) assistência social: responsável pela ajuda às pessoas necessitadas da sociedade. Compõem atualmente o departamento de assistência social da igreja o Lar Betânia, Bom Samaritano e o Centro de Recuperação de Viciados;
- d) casais: responsável pelos eventos e cursos para casados;
- e) círculo de oração: departamento das senhoras;
- f) comunicação: responsável pelos programas de rádio e televisão da igreja;
- g) ensino: responsável pela escola teológica e pela escola bíblica dominical;
- h) infantil: responsável pelas crianças da igreja, pelo culto infantil e pela preparação e capacitação de professores;
- i) jovens: é o departamento responsável pela juventude da igreja;
- j) missionário: responsável pelo envio de missionários às cidades e países necessitados como Cuba, Guiné Bissau, Moçambique e outros;
- k) terceira idade: responsável pelos idosos e aposentados da igreja.

5.1 MEMBROS DA IEADB

É declarado como membro da igreja a pessoa que é batizada na igreja, ou que apresenta carta de mudança vindo de outra igreja evangélica ou vindo de uma IEAD de outra cidade, mediante a avaliação do dirigente da congregação e posteriormente do pastor presidente da igreja.

Quando a pessoa torna-se membro da igreja são coletados diversos dados e informações deste novo membro para efetuação do seu cadastro no rol de membros da igreja.

Alguns tipos de membro existentes atualmente na IEADB são:

- a) pastores: responsáveis pela igreja;
- b) presbíteros, diáconos e auxiliares: são os cooperadores da igreja, ajudam os pastores na administração da igreja e geralmente dirigem algum departamento ou congregação da igreja;
- c) jovens, irmãos, irmãs e adolescentes: são os membros que freqüentam a igreja e geralmente não possuem nenhum cargo administrativo na igreja.

Algumas das possíveis situações de um membro da IEADB:

- a) ativo: o membro é considerado ativo quando está em comunhão com a igreja;
- b) disciplinado: o membro é considerado disciplinado quando alguma atitude de sua parte não é conveniente aos ensinamentos bíblicos seguidos pela igreja, porém a disciplina é apenas um período de tempo;
- c) excluído: quando o membro pede desligamento por completo da igreja, ou quando toma uma atitude totalmente adversa do ensinado na igreja e não quer e nem deseja mais ter comunhão com a igreja;
- d) falecido: após a sua morte o membro é considerado falecido, porém o registro continua na relação de membros, para que não se perca o seu histórico;
- e) mudou-se: muitos membros mudam-se de cidades, mas o seu registro continua na relação de membros alterando apenas a situação de “ativo” para “mudou-se”;

5.2 ENTRADAS DA IEADB

As entradas são as verbas da igreja que podem ser coletas de culto, ofertas, ou contribuições voluntárias de um membro da igreja ou alguma pessoa que mesmo não tendo vínculo algum com a igreja deseja fazer algum tipo de doação. Esta doação pode ser efetuada em qualquer congregação da igreja e pode ser atribuída a qualquer departamento. Por exemplo: O membro João da congregação Garcia efetuou uma contribuição de R\$ 25,00 para auxílio e manutenção dos missionários que estão em Cuba, logo essa entrada será atribuída ao departamento missionário e à congregação Garcia.

Todas as entradas da congregação são levadas até a congregação Sede onde fica localizada a tesouraria central da IEADB e então, é realizada a contabilidade destas contribuições e após este processo é enviado o relatório de volta para a congregação com o

nome de todos os contribuintes da congregação (Figura 6). Estes relatórios são entregues aos dirigentes da congregação em uma reunião geral entre dirigentes das congregações e a diretoria da igreja, que acontecem mensalmente na congregação Sede.

Após a verificação dos relatórios pelos dirigentes, os relatórios são expostos na congregação para o membro visualizar se realmente sua contribuição foi digitada corretamente no sistema.

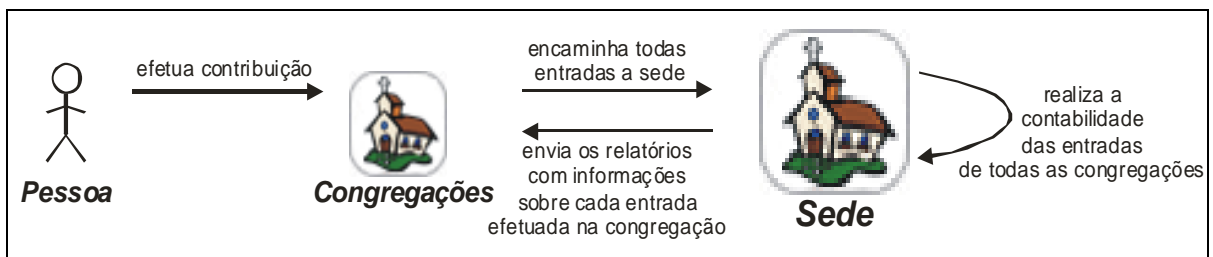


Figura 6 – Visualização do processo de entradas

5.3 SAÍDAS DA IEADB

As saídas são as despesas que a igreja possui como pagamento de água, luz, telefone, energia elétrica, zeladores, material de construção e outros. Estas despesas são pagas em sua maioria pela tesouraria central da igreja localizada na congregação Sede, que recebe as faturas das congregações e efetua o pagamento.

Assim como as entradas, as saídas podem ser efetuadas em qualquer congregação e podem ser atribuídas a qualquer departamento. Por exemplo: a congregação Itoupava Central adquiriu lápis de colorir para as crianças utilizarem durante o culto infantil, logo esta saída será atribuída ao departamento infantil e à congregação Itoupava Central.

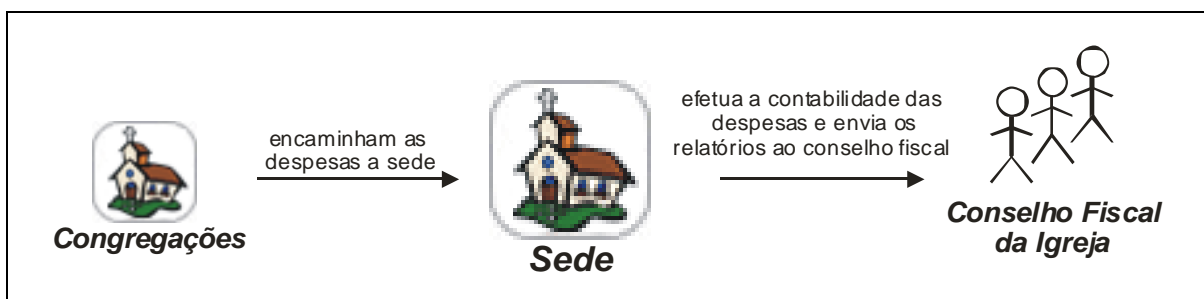


Figura 7 – Visualização do processo de saídas

5.4 PROBLEMA ATUAL

O grande problema é a intensa transação de informações entre todas as congregações e a congregação Sede onde fica localizada a secretaria e a tesouraria central da IEADB, pois todas as entradas e todas as saídas (despesas) têm que ser levadas até a congregação Sede, para serem conferidas, separadas, digitadas e organizadas para a confecção dos relatórios.

O problema não está associado só aos cadastros de entradas e saídas, pois qualquer alteração no cadastro de membros tem que ser efetuado na secretaria central. Se em um relatório o nome de um determinado membro estiver errado, ao perceber o erro, o membro avisa o secretário da congregação. O secretário da congregação então é obrigado a ir até a secretária central da igreja com o memorando de comunicação interna para que o nome deste membro seja devidamente alterado. Este tipo de problema gera burocracia, gastos adicionais, demora dos relatórios e insatisfação a todas as pessoas envolvidas neste processo.

O processo que ocorre atualmente na IEADB é centralizado totalmente na congregação Sede ficando a cargo das demais congregações somente o recebimento do relatório.

Este tipo de relacionamento intenso que existe entre as congregações e a sede não traz comodidade, independência e agilidade no tratamento das informações de cada congregação, muito pelo contrário, gera uma sobrecarga aos funcionários da secretaria e tesouraria central da igreja que em muitos casos se preocupam e gastam muito tempo com detalhes que poderiam ser resolvidos na própria congregação sem ter que passar pela sede.

Para algumas questões serem resolvidas na própria congregação é necessário no primeiro momento descentralizar parcialmente alguns processos, a fim de que as informações das congregações cheguem até a sede de forma que estejam parcialmente ou totalmente processadas, isto facilitará e agilizará todo o trabalho efetuado na secretaria e tesouraria central da igreja sede.

A visualização da transação de informações entre a sede e as demais congregações é visualizada parcialmente na Figura 8:

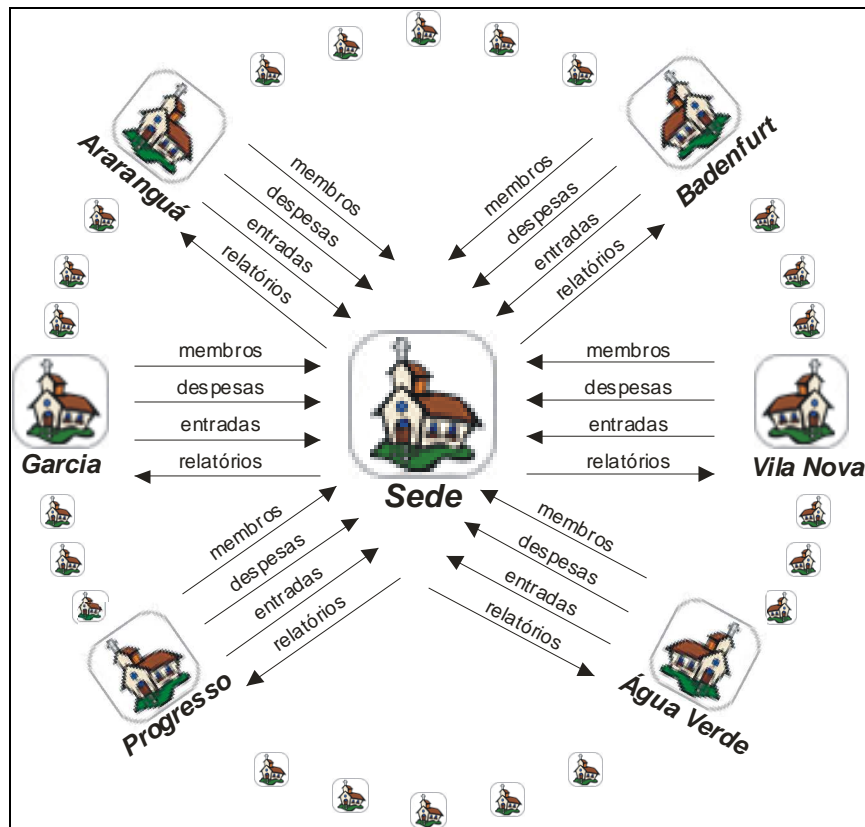


Figura 8 – Visualização parcial da transação de informações na IEADB

6 TRABALHOS CORRELATOS

Foram analisados dois sistemas para administração de igrejas já existentes no mercado e oferecido às igrejas evangélicas, o ROL-BL e o GESIG (Gestão de Igrejas).

6.1 ROL-BL

O ROL-BL é o sistema utilizado atualmente pela IEADB. Foi criado no ano de 1997 pela LAM Informática sob encomenda da IEAD de Chapecó-SC e desde então foi atualizado diversas vezes para se adaptar às mudanças da igreja, e para se adaptar às mudanças desejadas pelos usuários (administradores da igreja). O ROL-BL foi desenvolvido em Clipper e utiliza-se de bases DBF.

Algumas vantagens do ROL-BL:

- a) apresenta bom desempenho (rapidez) nas consultas efetuadas;
- b) facilita a digitação dos usuários no lançamento de dados no sistema;
- c) a utilização do ROL-BL não compromete o desempenho do sistema operacional.

Algumas desvantagens observadas no ROL-BL:

- a) não existe integridade de informações em alguns lançamentos de dados, pois é possível lançar uma entrada em uma determinada congregação, atribuindo a esta mesma entrada um setor diferente do qual a congregação pertence. Este problema pode acontecer no lançamento de saídas e no cadastro de membros também. É possível colocar uma irmã com o sexo masculino ou um irmão com o sexo feminino. Atributos como cidade e bairro não estão normalizados. Estes são alguns dos problemas de consistência de dados encontrados no ROL-BL sendo que estes problemas podem gerar confusão e muito transtorno na geração dos relatórios que são emitidos para as congregações;
- b) a interface gráfica (Figura 9 e Figura 10) é ruim pois o sistema é em MS-DOS o que dificulta a “navegação” pelo sistema, por isso é necessário possuir algum tempo de experiência e prática no sistema para poder encontrar as informações desejadas, ou seja, o ROL-BL não é de uso intuitivo;
- c) a segurança também é um ponto fraco do ROL-BL, pois é possível manipular as informações através do próprio Microsoft Excel, editando os arquivos DBF;

- d) é centralizado pois todas as informações tem que passar pelo computador onde o sistema está instalado.



Figura 9 – Visualização da listagem de membro no ROL-BL

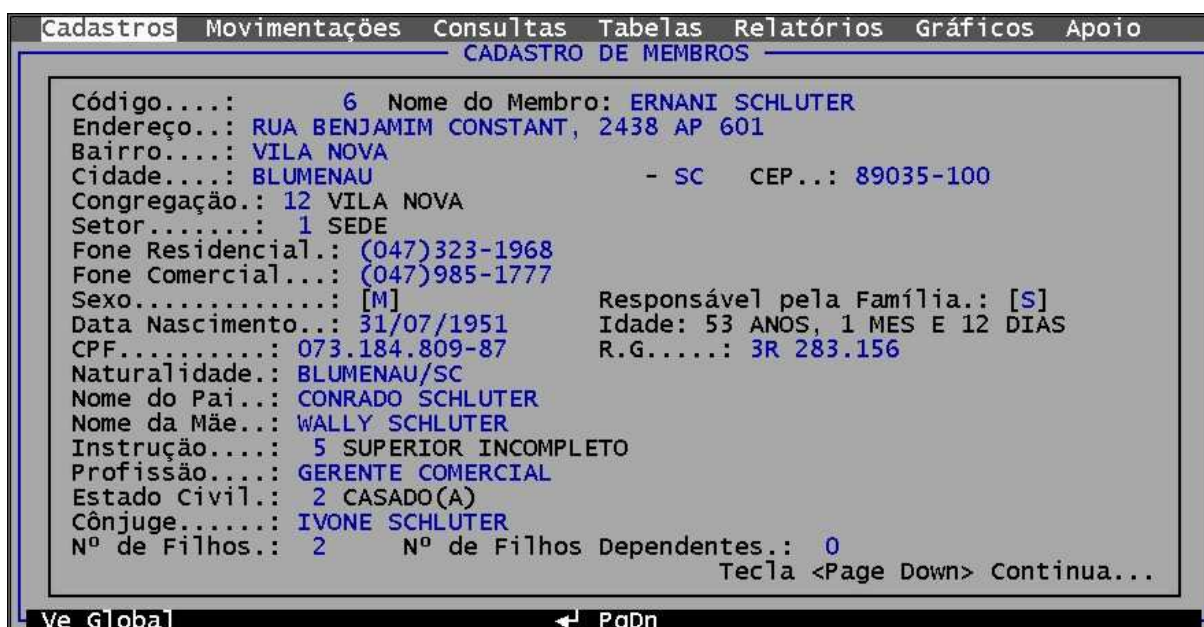


Figura 10 – Visualização do cadastro de membro no ROL-BL

O ROL-BL é utilizado atualmente nas IEADs de Blumenau, Chapecó, Lages, Fraiburgo e em Timbó Grande.

Ainda hoje este sistema é distribuído e atualizado pela LAM Informática, rua José Bonifácio, 352D, bairro Jardim América em Chapecó-SC.

6.2 GESIG

O GESIG (Figura 11) é um sistema atual em relação ao ROL-BL, e segundo seus desenvolvedores apresentam as seguintes vantagens:

- a) padrões de interface gráfica Windows 95/98;
- b) barras de atalho e ferramentas configuradas pelo usuário;
- c) controle de acesso através de senhas;
- d) compatibilidade com todos os bancos de dados do mercado;
- e) versões monousuário e multiusuário;
- f) fácil aprendizado e utilização;
- g) faz integridade antes do lançamento dos dados no banco de dados.

The screenshot displays the 'HDS Sistemas - Gestão de Igrejas' application window. The title bar reads 'HDS Sistemas Ltda - Gestão de Igrejas'. The menu bar includes 'Arquivo', 'Lançamentos', 'Consultas', 'Relatório', 'Opções', and 'Janela ?'. The main window title is 'HDS Sistemas - Gestão de Igrejas - Dados Pessoais'. Below the title bar is a toolbar with navigation and action icons. The main area contains a form for 'Identificação Pessoal' with a dropdown menu set to '1'. The form is divided into several sections: 'Dados Pessoais', 'Características Físicas', 'Endereço', 'Atividade Profissional', 'Membros', 'Entrevista', and 'Dependentes'. The 'Dados Pessoais' section includes fields for 'Nome' (JOHN DUMBAR), 'Data de Nascimento' (11/11/68), 'Local de Nascimento' (CURITIBA), and 'Categoria Pessoa' (PASTOR TITULAR). The 'Filiação' section includes 'Nome do Pai' (MARIO DUMBAR) and 'Nome da Mãe' (MARIANA DUMBAR). The 'Complemento' section includes 'Estado Civil' (C), 'Grau de Instrução' (8), 'Naturalidade' (PR), and 'Nacionalidade' (10). The 'Identificação' section includes 'Identidade' (2.195.999-6) and 'CPF' (555.555.666). The 'Sexo' section has radio buttons for 'Masculino' (selected) and 'Feminino'. The 'Cônjuge' section includes 'Nome' (JOANA DUMBAR) and 'Data Nasc' (11/09/80). The 'Título Eleitoral' section includes 'Cidade' (JOINVILLE), 'Número' (1234), and 'Zona' (12). At the bottom of the form, it says 'Última Alteração: HDS-26/03/98-16:14:22'. Below the form is a toolbar with buttons for 'Incluir', 'Atualizar', 'Excluir', and 'Ignorar'.

Figura 11 – Visualização dos dados pessoais do membro no GESIG

Mesmo sendo um sistema atual e com interface Windows o GESIG é centralizado pois não possui níveis de usuários e só pode ser utilizado em uma LAN.

O GESIG é distribuído pela Eclésia Soluções e Sistemas, rua Almirante Jaceguay, 3704, bairro Costa e Silva em Joinville-SC.

7 DESENVOLVIMENTO DO TRABALHO

Este capítulo trata do desenvolvimento do protótipo do sistema web. São apresentados os requisitos do sistema, a especificação e a implementação.

7.1 REQUISITOS PRINCIPAIS DO PROBLEMA

O sistema deverá auxiliar os administradores na administração da igreja sede e todas as congregações, devendo também diminuir a burocracia entre todas as congregações que estão ligadas à igreja sede (escritório central).

Na elicitação e na pesquisa de requisitos do sistema percebeu-se a necessidade de no primeiro momento dividir em três níveis de administradores, em função da utilização da web.

O nível um e principal será utilizado pelo administrador global da igreja que tem uma visão de todo o sistema. O administrador global tem permissão para efetuar qualquer tipo de cadastro, alteração e exclusão em qualquer congregação da igreja e também cadastra os demais administradores que poderão fazer uso do sistema.

O nível dois é utilizado pelos administradores que trabalham no escritório central da igreja sede, chamado de administrador central. O administrador central tem permissão para efetuar qualquer tipo de cadastro, alteração e exclusão em qualquer congregação da igreja, mas não possui permissão para cadastrar novos administradores no sistema.

O nível três é utilizado pelos administradores que pertencem a alguma congregação da igreja, chamado de administrador congregacional. O administrador congregacional tem permissão para efetuar cadastros, exclusão ou e alteração de membros, entradas e saídas pertinentes a congregação a qual ele pertence, não podendo visualizar nenhum conteúdo das demais congregações.

Como características importantes a serem implementadas no sistema encontra-se também a segurança através de senhas e nível dos usuários, a utilização de software livre e a compatibilidade com o navegador mais utilizado no mercado, o Internet Explorer.

Para facilitar o entendimento dos requisitos do sistema desenvolvido, tem-se a visão dos administradores e dos cadastros principais que foram implementados no sistema (Figura 12).

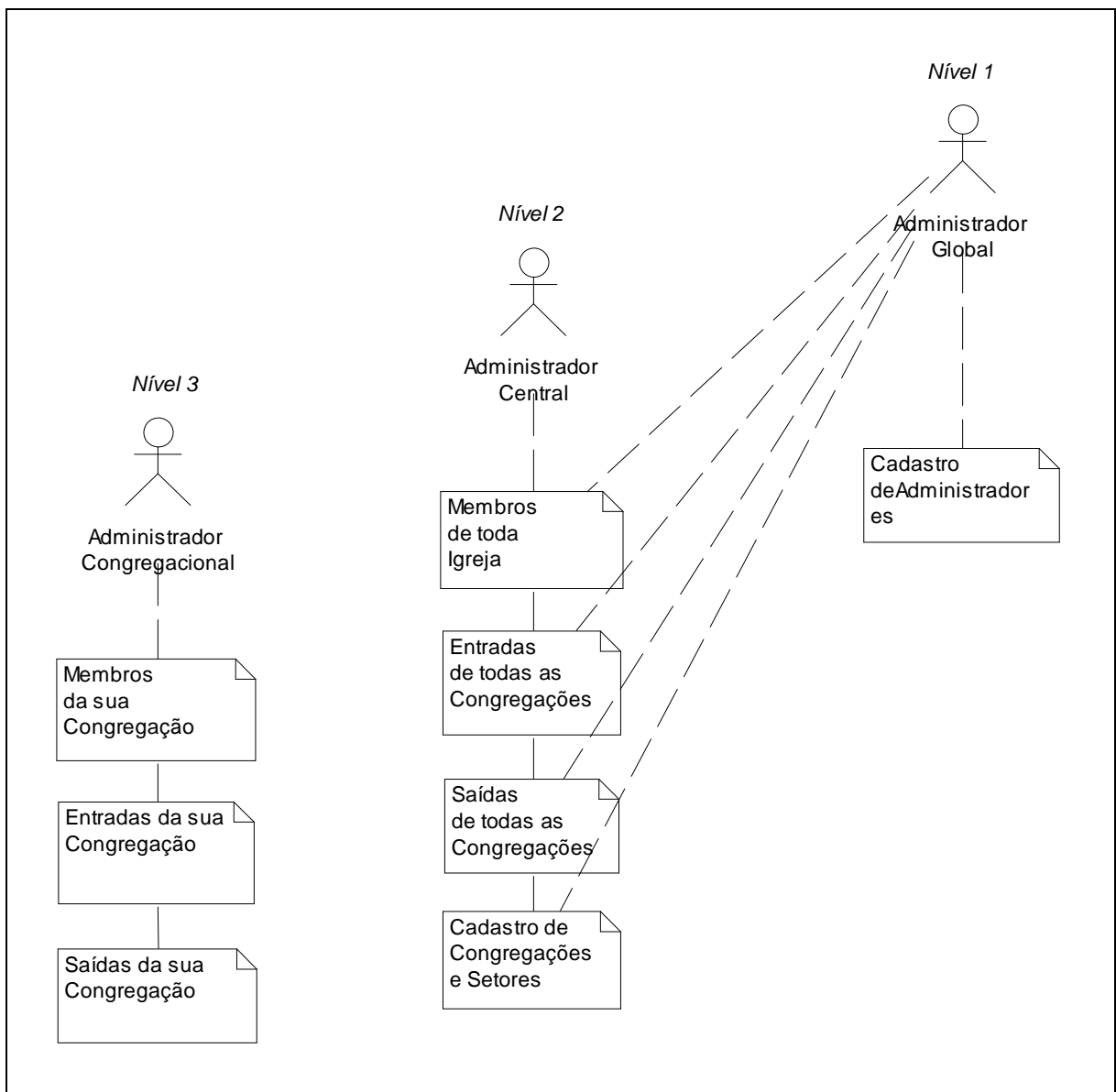


Figura 12 – Visualização dos usuários (administradores) envolvidos no sistema

7.2 ESPECIFICAÇÃO

Para a especificação do sistema implementado utilizou-se a ferramenta Rational Rose 2002. A seguir são apresentados os diagramas de casos de uso, a partir dos quais se pode ter uma visão macro das funcionalidades do sistema. Conforme Bezerra (2002, p.57), os diagramas de caso de uso (DCU) correspondem a uma visão externa do sistema e representa graficamente os atores envolvidos, casos de uso e relacionamento entre esses elementos. O diagrama de casos de uso tem o objetivo de ilustrar em um nível alto de abstração quais elementos externos interagem com que funcionalidades do sistema.

7.2.1 NÍVEL 1 E NÍVEL 2

As diferenças do nível 1 (administrador global) e do nível 2 (administrador central) são pequenas e por esta razão serão visualizadas na mesma figura. O nível 1 e 2 são os administradores que trabalham no escritório central da igreja e tem permissão para fazer cadastros e alterações em todas as congregações da igreja, pois estes administradores serão responsáveis pelas informações de todas as congregações e efetuarão os serviços daquelas congregações que não possuem um administrador congregacional. Na Figura 13 é mostrado o diagrama de casos de uso do administrador global e do administrador central e os principais casos de uso primários pertinentes a cada um.



Figura 13 – Diagrama de casos de uso do administrador central e do administrador global

A seguir são descritos brevemente os principais casos de uso destes níveis:

- a) cadastrar administradores (usuários): este processo é permitido somente ao administrador global, que de acordo com a necessidade cadastra novos administradores para trabalhar no sistema e também decide o nível de cada novo administrador;
- b) cadastrar membros: os administradores do nível 1 e 2 podem cadastrar membros em qualquer congregação e fazer alterações no cadastro desses membros.
- c) cadastrar entradas: as entradas que são contribuições dos membros à igreja, podem ser cadastradas pelos administradores do nível 1 e 2 sendo atribuída à congregação onde a contribuição foi efetuada..
- d) cadastrar saídas: as saídas que são as despesas, serão lançadas pelos administradores do nível 1 e 2 sendo atribuídas a congregação que efetuou a despesa.
- e) cadastrar congregações: os administradores do nível 1 e 2 tem permissão para criar novas congregações, alterar e excluir as congregações.
- f) cadastrar setores: os administradores do nível 1 e 2 também tem permissão para criar, alterar e excluir novos setores.
- g) gerar relatórios: os relatórios de todas as congregações podem ser visualizados e impresso pelos administradores do nível 1 e 2, para que possam entregar as congregações que não possuem um administrador congregacional;
- h) efetuar mineração de dados: para auxiliar os administradores da igreja na tomada de decisão é verificado o andamento da congregação em um determinado período de tempo através de técnicas de *data mining*.

7.2.2 NÍVEL 3

O nível 3 é o administrador congregacional, cujo diagrama de casos de uso primários é mostrado na Figura 14.

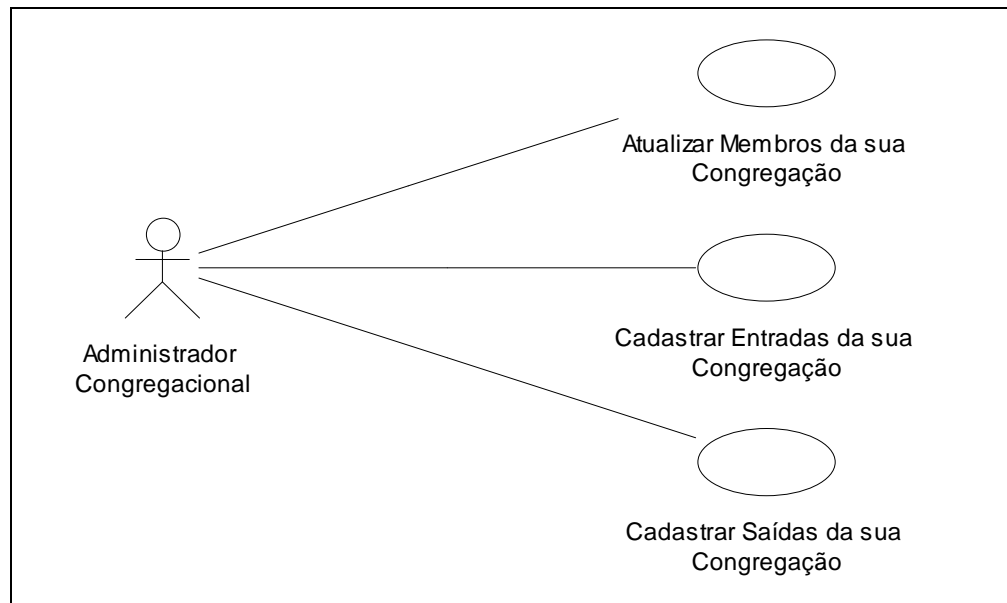


Figura 14 – Diagrama de casos de uso do administrador congregacional

A seguir são descritos brevemente os principais casos de uso deste nível:

- a) atualizar membros da sua congregação: o administrador congregacional poderá somente atualizar os membros da sua própria congregação, pois para um novo membro ser cadastrado o pastor geral deve estar ciente, por isso esta tarefa ainda deve passar pelo escritório central.
- b) cadastrar entradas da sua congregação: o administrador congregacional poderá lançar no sistema e fazer alterações nas entradas da sua própria congregação, mas não terá permissão para visualizar nenhuma outra congregação a não ser a sua.
- c) cadastrar saídas da sua congregação: o administrador congregacional poderá também cadastrar as saídas da sua própria congregação.

Para uma melhor compreensão do armazenamento dos dados, é apresentado na Figura 15 o Diagrama Entidade Relacionamento (DER) lógico do sistema, com todas as entidades que formam a base de dados e seus relacionamentos. Conforme Heuser (2000, p.11) o DER é a técnica de modelagem de dados mais difundida e utilizada na captura das necessidades da organização em termos de armazenamento de dados de forma independente de

implementação, ou seja, a partir do DER obtêm-se uma descrição abstrata, independente de implementação em computador, dos dados que serão armazenado no banco de dados.

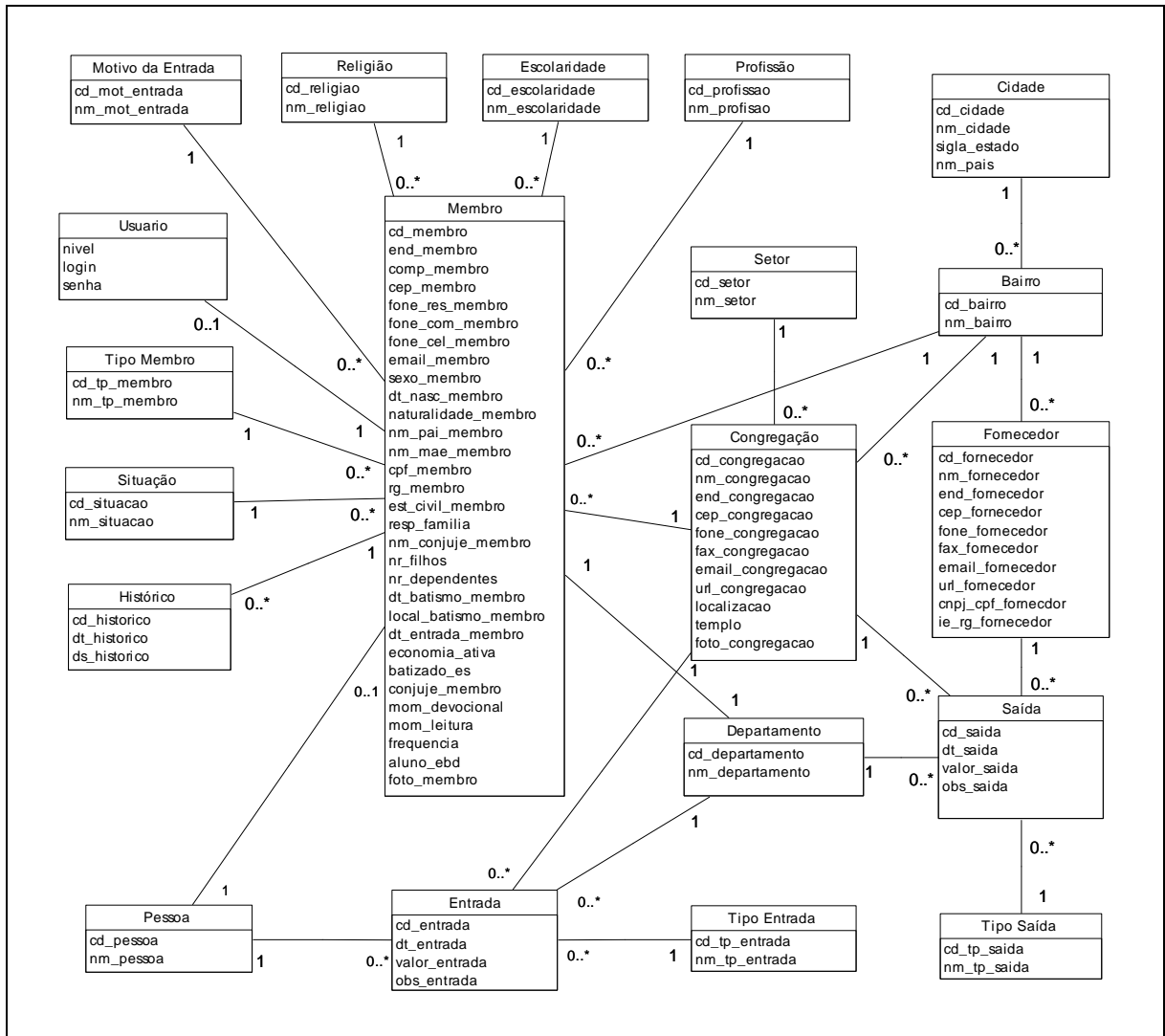


Figura 15 – Diagrama entidade relacionamento lógico

Para complementar o DER lógico no Quadro 2 é representado o dicionário de dados físico detalhando todos os atributos de cada entidade.

BAIRRO			
Nome	Descrição	Tipo	Chave primária
cd_bairro	Código do bairro	int(11)	sim
nm_bairro	Nome do bairro	varchar(50)	
cd_cidade	Código da cidade	int(11)	

CIDADE			
Nome	Descrição	Tipo	Chave primária
cd_cidade	Código da cidade	int(11)	sim
nm_cidade	Nome da cidade	varchar(50)	
sigla_estado	Sigla do estado	char(2)	

nm_pais	Nome do país	varchar(50)	
---------	--------------	-------------	--

CONGREGAÇÃO			
Nome	Descrição	Tipo	Chave primária
cd_congregacao	Código da congregação	int(11)	sim
nm_congregacao	Nome da congregação	varchar(50)	
cd_setor	Código do setor	int(11)	
end_congregacao	Endereço da congregação	varchar(100)	
cep_congregacao	Cep da congregação	varchar(9)	
cd_bairro	Código do bairro da congregação	int(11)	
cd_membro	Código do membro responsável	int(11)	
fone_congregacao	Telefone da congregação	varchar(30)	
fax_congregacao	Fax da congregação	varchar(30)	
email_congregacao	E-mail da congregação	varchar(50)	
url_congregacao	URL da congregação	varchar(50)	
dm_localização	Localização da congregação (boa ou ruim)	varchar(10)	
dm_templo	Templo construído (sim ou não)	varchar(10)	
foto_congregação	Foto da congregação	imagem jpeg	

DEPARTAMENTO			
Nome	Descrição	Tipo	Chave primária
cd_departamento	Código do departamento	int(11)	sim
nm_departamento	Nome do departamento	varchar(50)	
cd_membro	Código do membro responsável	int(11)	

ENTRADA			
Nome	Descrição	Tipo	Chave primária
cd_entrada	Número do recibo de entrada	varchar(20)	sim
dt_entrada	Data da entrada	int(8)	sim
cd_departamento	Código do departamento	int(11)	
cd_tp_entrada	Código do tipo de entrada	int(11)	
cd_pessoa	Código da pessoa que efetuou a entrada	int(11)	
cd_congregacao	Código da congregação	int(11)	
obs_entrada	Observação para a entrada	varchar(255)	
valor_entrada	Valor da entrada	decimal(10,2)	

ESCOLARIDADE			
Nome	Descrição	Tipo	Chave primária
cd_escolaridade	Código da escolaridade	int(11)	sim
nm_escolaridade	Nome da escolaridade	varchar(50)	

FORNECEDOR			
Nome	Descrição	Tipo	Chave primária
cd_fornecedro	Código do fornecedor	int(11)	sim
nm_fornecedor	Nome do fornecedor	varchar(100)	
end_fornecedor	Endereço do fornecedor	varchar(100)	
cep_fornecedor	Cep do fornecedor	varchar(9)	
cd_bairro	Código do bairro do fornecedor	int(11)	
fone_fornecedor	Telefone do fornecedor	varchar(30)	
fax_fornecedor	Fax do fornecedor	varchar(30)	
email_fornecedor	E-mail do fornecedor	varchar(50)	
url_fornecedor	URL do fornecedor	varchar(50)	
cnpj_cpf_fornecedor	CNPJ ou CPF do fornecedor	varchar(14)	

ie_rg_fornecedor	IE ou RG do fornecedor	varchar(20)	
MEMBRO			
Nome	Descrição	Tipo	Chave primária
cd_membro	Código do membro	int(11)	sim
end_membro	Endereço do membro	varchar(100)	
comp_membro	Complemento do endereço do membro	varchar(100)	
cep_membro	Cep do membro	varchar(9)	
cd_bairro	Código do bairro do membro	int(11)	
cd_congregacao	Código da congregação do membro	int(11)	
fone_res_membro	Telefone residencial do membro	varchar(30)	
fone_com_membro	Telefone comercial do membro	varchar(30)	
fone_cel_membro	Telefone celular do membro	varchar(30)	
email_membro	E-mail do membro	varchar(50)	
sexo_membro	Sexo do membro	char(1)	
dt_nas_membro	Data de nascimento do membro	int(8)	
naturalidade	Código da cidade onde o membro nasceu	int(11)	
nm_pai_membro	Nome do pai	varchar(100)	
nm_mae_membro	Nome da mãe	varchar(100)	
cpf_membro	CPF do membro	varchar(14)	
rg_membro	RG do membro	varchar(20)	
est_civil	Estado civil do membro	varchar(50)	
nm_conjuge	Nome do cônjuge do membro	varchar(100)	
resp_familia	Responsável pela família	int(1)	
nr_filhos	Número de filhos	int(2)	
nr_dependentes	Número de dependentes	int(2)	
cd_escolaridade	Código da escolaridade do membro	int(11)	
cd_profissao	Código da profissão do membro	int(11)	
cd_tp_membro	Código do tipo de membro	int(11)	
dt_ordenacao	Data de ordenação	int(8)	
nm_ordenador	Nome do ordenador	varchar(100)	
cd_religiao	Código da religião anterior do membro	int(11)	
cd_mot_entrada	Código do motivo de entrada do membro	int(11)	
dt_batismo	Data de batismo do membro	int(8)	
local_batismo	Código do local do batismo	int(11)	
admitido_por	Nome da pessoa que admitiu o membro	varchar(100)	
cd_situacao	Código da situação atual do membro	int(11)	
economia_ativa	Membro economicamente ativo	int(1)	
batizado_es	Batizado no Espírito Santo	int(1)	
cônjuge_membro	Cônjuge é membro	int(1)	
mom_devocional	Membro tem momento de devocional	int(1)	
mom_leitura	Membro tem momento de leitura	int(1)	
frequência	Membro freqüente assiduamente a igreja	int(1)	
aluno_ebd	Membro é aluno da Escola Bíblica	int(1)	
foto_membro	Foto do membro da igreja	imagem jpeg	
MOTIVO_ENTRADA			
Nome	Descrição	Tipo	Chave primária
cd_mot_entrada	Código do motivo de entrada	int(11)	sim
nm_mot_entrada	Nome do motivo de entrada	varchar(50)	
PESSOA			
Nome	Descrição	Tipo	Chave primária

cd_pessoa	Código da pessoa	int(11)	sim
nm_pessoa	Nome da pessoa	varchar(100)	

PROFISSÃO			
Nome	Descrição	Tipo	Chave primária
cd_profissao	Código da profissão	int(11)	sim
nm_profissao	Nome da profissão	varchar(50)	

RELIGIAO			
Nome	Descrição	Tipo	Chave primária
cd_religio	Código da religião anterior	int(11)	sim
nm_religio	Nome da religião	varchar(50)	

SAIDA			
Nome	Descrição	Tipo	Chave primária
cd_saida	Número do recibo de saída	varchar(20)	sim
dt_saida	Data da entrada	int(8)	sim
cd_departamento	Código do departamento	int(11)	
cd_congregacao	Código da congregação	int(11)	
cd_tp_saida	Código do tipo de saída	int(11)	
obs_saida	Observação para a saída	varchar(255)	
cd_fornecedor	Código do fornecedor	int(11)	
valor_saida	Valor da saída	decimal(10,2)	

SETOR			
Nome	Descrição	Tipo	Chave primária
cd_setor	Código do setor	int(11)	sim
nm_setor	Nome do setor	varchar(50)	

SITUACAO			
Nome	Descrição	Tipo	Chave primária
cd_situacao	Código da situação atual	int(11)	sim
nm_situacao	Nome da situação	varchar(50)	

TIPO_ENTRADA			
Nome	Descrição	Tipo	Chave primária
cd_tp_entrada	Código do tipo de entrada	int(11)	sim
nm_tp_entrada	Nome do tipo de entrada	varchar(50)	

TIPO_MEMBRO			
Nome	Descrição	Tipo	Chave primária
cd_tp_membro	Código do tipo de membro	int(11)	sim
nm_tp_membro	Nome do tipo de membro	varchar(50)	

TIPO_SAIDA			
Nome	Descrição	Tipo	Chave primária
cd_tp_saida	Código do tipo de saída	int(11)	sim
nm_tp_saida	Nome do tipo de saída	varchar(50)	

USUÁRIO			
Nome	Descrição	Tipo	Chave primária
cd_membro	Código do membro que é usuário	int(11)	sim
nível	Nível do usuário	int(1)	

login	Login do usuário	varchar(12)	
senha	Senha do usuário	varchar(12)	

Quadro 2 – Dicionário de dados do sistema

7.3 A MINERAÇÃO DE DADOS DO SISTEMA

O objetivo da mineração de dados no sistema é classificar as congregações de acordo com o seu andamento para auxiliar os administradores da igreja fazendo uso da técnica árvore de decisão já citada no capítulo 4.2 para implementar a mineração de dados no sistema, pois árvores de decisão é uma das mais utilizadas técnicas de DM e de simples compreensão.

A coleção de dados foi desenvolvida a partir de informações fornecidas pelos responsáveis pela administração da IEADB, onde o atributo alvo é o andamento da congregação e os outros atributos são os que influenciam no andamento da congregação.

O atributo alvo servirá para indicar o andamento da congregação aos administradores da igreja e, baseada nesta informação, tomar as decisões necessárias para rever o que é necessário fazer quando o andamento da congregação é igual a regular ou ruim.

A coleção de dados representada pela Tabela 1 serve para visualizar as situações que acontecem nas congregações da IEADB em relação aos seguintes atributos:

- a) número de membros da congregação;
 - acima de 150 = grande;
 - entre 75 e 150 = média;
 - abaixo de 75 = pequena;
- b) número de contribuintes da congregação em relação ao número de membros;
 - maior que 65% = ótimo;
 - entre 45% e 65% = bom;
 - entre 25% e menor que 45% = ruim;
 - menor que 25% = péssimo;
- c) se a congregação possui templo construído;
 - possui templo construído = sim;
 - não possui templo construído = não;
- d) quanto a localização da congregação;
 - bem localizada = boa;
 - mal localizada = ruim.

Tabela 1 – Tabela de decisão do andamento da congregação

	Número de Membros	Número de Contribuintes	Templo Construído	Localização	Andamento
1	grande	ótimo	sim	boa	bem
2	grande	bom	sim	boa	regular
3	grande	ruim	sim	boa	mal
4	grande	péssimo	sim	boa	mal
5	média	ótimo	sim	boa	bem
6	média	bom	sim	boa	regular
7	média	ruim	sim	boa	mal
8	média	péssimo	sim	boa	mal
9	pequena	ótimo	sim	boa	regular
10	pequena	bom	sim	boa	regular
11	pequena	ruim	sim	boa	mal
12	pequena	péssimo	sim	boa	mal
13	grande	ótimo	não	boa	bem
14	grande	bom	não	boa	bem
15	grande	ruim	não	boa	mal
16	grande	péssimo	não	boa	mal
17	média	ótimo	não	boa	bem
18	média	bom	não	boa	bem
19	média	ruim	não	boa	mal
20	média	péssimo	não	boa	mal
21	pequena	ótimo	não	boa	bem
22	pequena	bom	não	boa	regular
23	pequena	ruim	não	boa	mal
24	pequena	péssimo	não	boa	mal
25	grande	ótimo	sim	ruim	bem
26	grande	bom	sim	ruim	bem
27	grande	ruim	sim	ruim	regular
28	grande	péssimo	sim	ruim	mal
29	média	ótimo	sim	ruim	bem
30	média	bom	sim	ruim	bem
31	média	ruim	sim	ruim	regular
32	média	péssimo	sim	ruim	mal
33	pequena	ótimo	sim	ruim	regular
34	pequena	bom	sim	ruim	regular
35	pequena	ruim	sim	ruim	mal
36	pequena	péssimo	sim	ruim	mal
37	grande	ótimo	não	ruim	bem
38	grande	bom	não	ruim	bem
39	grande	ruim	não	ruim	bem
40	grande	péssimo	não	ruim	mal
41	média	ótimo	não	ruim	bem
42	média	bom	não	ruim	bem
43	média	ruim	não	ruim	regular
44	média	péssimo	não	ruim	mal
45	pequena	ótimo	não	ruim	bem
46	pequena	bom	não	ruim	bem
47	pequena	ruim	não	ruim	regular
48	pequena	péssimo	não	ruim	mal

De acordo com Quinlan (1993) o processo de formação da árvore de decisão começa com a definição de que atributo será o nó inicial da árvore (nó raiz). Para isso deve se calcular a entropia que é o conceito utilizado para determinar o fator de incidência de cada atributo em relação ao atributo alvo, neste caso o andamento da congregação.

O processo de cálculo da entropia (capítulo 4.2) começa com a seleção distinta dos valores do atributo alvo. Conforme a tabela 1 os valores distintos para o atributo decisivo são (bem, regular, mal). Então calcula-se a quantidade de vezes que cada um desses valores ocorrem dentro da coleção de dados (Quadro 3).

Quantidade de ocorrências: bem = 18, regular = 11, mal = 19

Quantidade total de ocorrências: 48

$$\text{Entropia}(S) = \sum_{-1} p(I) \log_2 p(I)$$

I = Quantidade de ocorrências para valor distinto do atributo dividido pela quantidade total de ocorrências.

S = Coleção de dados da tabela 1

$$P1 = 18/48 \text{ (bem)}$$

$$P2 = 11/48 \text{ (regular)}$$

$$P3 = 19/48 \text{ (mal)}$$

$$\text{Entropia}(S) = -(18/48) * \log_2(18/48) - (11/48) * \log_2(11/48) - (19/48) * \log_2(19/48)$$

$$\text{Entropia}(S) = 1,54698$$

Quadro 3 – Cálculo de entropia do atributo alvo

Depois de calculado a entropia do atributo alvo deve-se calcular o valor de ganho para cada atributo não decisivo, e então o atributo não decisivo que possuir maior ganho será considerado o atributo inicial da árvore, fazendo uso da fórmula $\text{Ganho}(S,A) = \text{Entropia}(S) - \sum ((|S_v|) / |S|) * \text{Entropia}(S_v)$.

Para calcular o valor de ganho deve-se selecionar os valores distintos de cada atributo não decisivo e contar a quantidade de vezes que cada um desses atributos ocorrem em relação ao atributo alvo, calculando-se também a sua entropia, conforme demonstração no Quadro 4, Quadro 5, Quadro 6 e Quadro 7.

Valores para número de membros da congregação:

(grande, média, pequena)

Quantidade de ocorrências: grande=16, média=16, pequena=16

Entropia de (grande) = $-(8/16) * \log_2(8/16) - (2/16) * \log_2(2/16) - (6/16) * \log_2(6/16)$

Entropia de (grande) = 1,405

Entropia de (média) = $-(7/16) * \log_2(7/16) - (3/16) * \log_2(3/16) - (6/16) * \log_2(6/16)$

Entropia de (média) = 1,505

Entropia de (pequena) = $-(3/16) * \log_2(3/16) - (6/16) * \log_2(6/16) - (7/16) * \log_2(7/16)$

Entropia de (pequena) = 1,505

Ganho do atributo número de membros:

$1,54698 - [(16/48) * 1,405] + [(16/48) * 1,505] + [(16/48) * 1,505] = 0,07531$

Quadro 4 – Cálculo de ganho do atributo número de membros da congregação

Valores para número de contribuintes da congregação:

(ótimo, bom, regular, péssimo)

Quantidade de ocorrências: ótimo=12, bom=12, ruim=12, péssimo=12

Entropia de (ótimo) = $-(10/12) * \log_2(10/12) - (2/12) * \log_2(2/12) - (0/12) * \log_2(0/12)$

Entropia de (ótimo) = 0,650

Entropia de (bom) = $-(7/12) * \log_2(7/12) - (5/12) * \log_2(5/12) - (0/12) * \log_2(0/12)$

Entropia de (bom) = 0,979

Entropia de (ruim) = $-(1/12) * \log_2(1/12) - (4/12) * \log_2(4/12) - (7/12) * \log_2(7/12)$

Entropia de (ruim) = 1,280

Entropia de (péssimo) = $-(0/12) * \log_2(0/12) - (0/12) * \log_2(0/12) - (12/12) * \log_2(12/12)$

Entropia de (péssimo) = 0

Ganho do atributo número de contribuintes:

$1,54698 - [(12/48) * 0,650] + [(12/48) * 0,979] + [(12/48) * 1,280] + 0 = 0,81973$

Quadro 5 – Cálculo de ganho do atributo número de contribuintes da congregação

Valores para templo construído:

(sim, não)

Quantidade de ocorrências: sim=24, não=24

Entropia (sim) = $-(6/24) * \log_2(6/24) - (8/24) * \log_2(8/24) - (10/24) * \log_2(10/24)$

Entropia (sim) = 1,554

Entropia (não) = $-(12/24) * \log_2(12/24) - (3/24) * \log_2(3/24) - (9/24) * \log_2(9/24)$

Entropia (não) = 1,405

Ganho do atributo templo construído:

$1,54698 - [(24/48) * 1,554] + [(24/48) * 1,405] = 0,06748$

Quadro 6 – Cálculo de ganho do atributo templo construído

Valores para localização:

(boa, ruim)

Quantidade de ocorrências: boa=24, ruim=24

Entropia (boa) = $-(7/24) * \log_2(7/24) - (5/24) * \log_2(5/24) - (12/24) * \log_2(12/24)$

Entropia (boa) = 1,489

Entropia (ruim) = $-(11/24) * \log_2(11/24) - (6/24) * \log_2(6/24) - (7/24) * \log_2(7/24)$

Entropia (ruim) = 1,534

Ganho do atributo localização:

$1,54698 - [(24/48) * 1,489] + [(24/48) * 1,534] = 0,03548$

Quadro 7 – Cálculo de ganho do atributo localização da congregação

Comparando os valores de ganho de cada atributo verifica-se que o atributo número de contribuintes possui o maior ganho, logo este atributo será o nó raiz da árvore. Depois de se dividir as ramificações a partir do nó raiz, deve-se calcular a entropia para cada uma das novas ramificações existentes para saber qual será o próximo nó da árvore. Para isto deve-se considerar cada sub-conjunto gerado pelo valor dos atributos do nó raiz.

Para abstrair os demais cálculos e para auxiliar no processo de construção da árvore de decisão foi utilizado a ferramenta WEKA (*Waikato Environment for Knowledge Analysis*) demonstrado e utilizado na disciplina de Inteligência Artificial na FURB. O pacote da WEKA é formado por um conjunto de implementações de algoritmos de diversas técnicas de mineração de dados. O pacote WEKA é um software de domínio público, e está disponível no seguinte endereço eletrônico: <http://www.cs.waikato.ac.nz/ml/weka/>

Após a tabela de decisão (Tabela 1) ser submetida à ferramenta WEKA obteve-se os seguintes resultados para o algoritmo ID3 (Figura 16) e o algoritmo C4.5 (Figura 17):

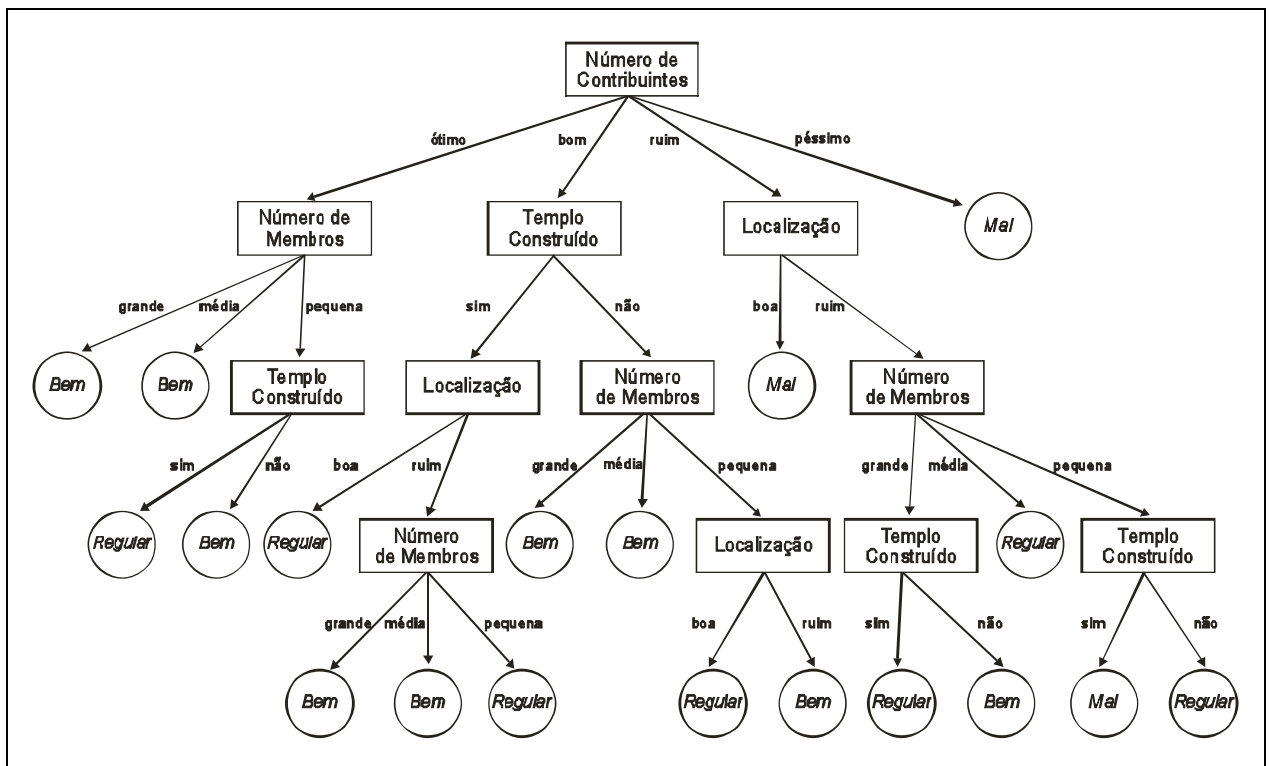


Figura 16 – Árvore de decisão gerada pelo algoritmo ID3 sem podagem de árvore

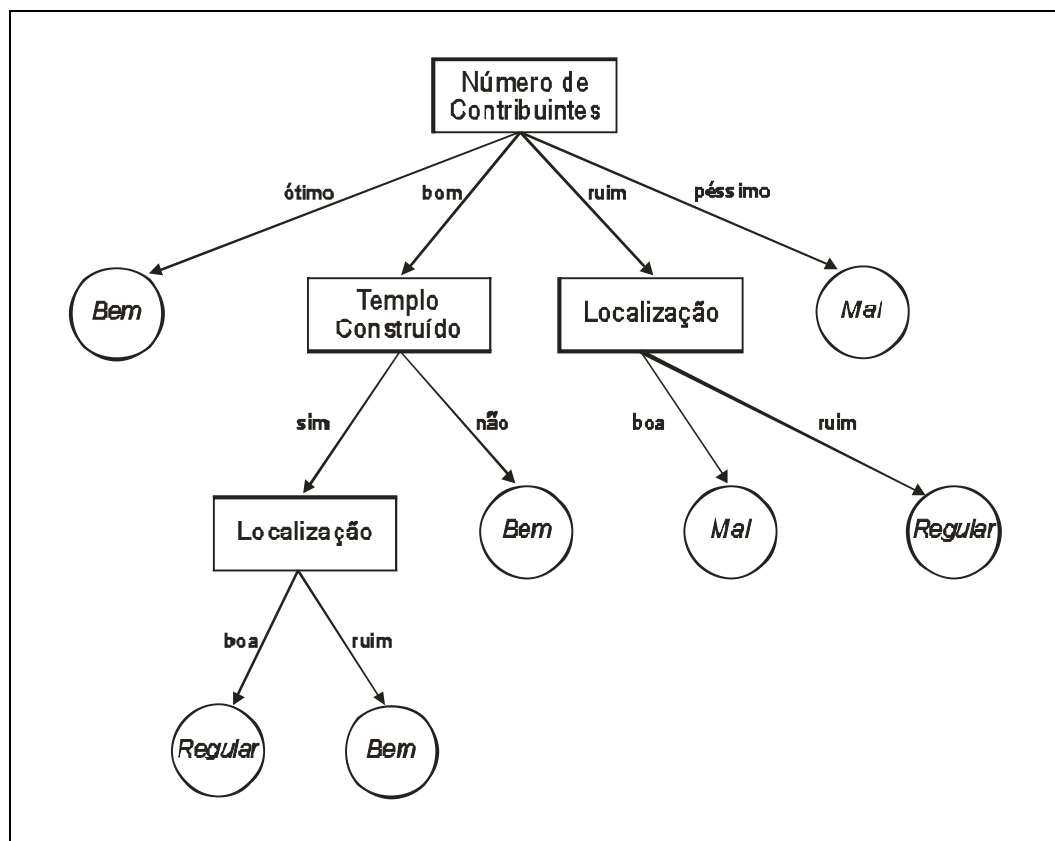


Figura 17 – Árvore de decisão gerada pelo algoritmo C4.5 com podagem de árvore

O algoritmo ID3 gerou uma árvore de tamanho igual a trinta e um (31), enquanto o algoritmo C4.5 gerou uma árvore de tamanho onze (11) pois efetua podagem de árvore, mas o algoritmo C4.5 apesar de possuir uma performance melhor apresenta também um total de seis (6) linhas “perdidas” da tabela de decisão, pois foram cortados da árvore, por possuírem pouquíssimas ocorrências. Nota-se, por exemplo, que nem todas as congregações com número de contribuintes igual a ótimo o andamento é igual a bem (linha 9 e 33 da tabela de decisão). Significa que fazendo uso do algoritmo C4.5 alguns resultados podem ser diferentes da tabela de decisão devido a exclusão de alguns nós da árvore de decisão.

Após a construção da árvore foram geradas as regras que verificarão o resultado do andamento da congregação. As regras para o algoritmo com podagem de árvore C4.5 são demonstradas no Quadro 3 em código PHP, assim como está implementado no sistema. Os valores para os atributos, número de membros, número de contribuintes, localização e templo construído são efetuados a partir de consultas ao banco de dados. Depois de efetuada estas consultas os valores são submetido às regras, que indicarão o andamento da congregação avaliada.

A ferramenta *WEKA* construiu a árvore de decisão para o algoritmo *id3* como visualizado parcialmente na Figura 18.

```
Classifier output
NmContribuintes = otimo
| NmMembros = grande: bem
| NmMembros = media: bem
| NmMembros = pequena
| | templo = sim: regular
| | templo = nao: bem
NmContribuintes = bom
| templo = sim
| | localização = boa: regular
| | localização = ruim
| | | NmMembros = grande: bem
| | | NmMembros = media: bem
| | | NmMembros = pequena: regular
| | templo = nao
| | | NmMembros = grande: bem
| | | NmMembros = media: bem
| | | NmMembros = pequena
| | | | localização = boa: regular
| | | | localização = ruim: bem
NmContribuintes = ruim
| localização = boa: mal
| localização = ruim
| | NmMembros = grande
| | | templo = sim: regular
```

Figura 18 – Árvore construída pela ferramenta *WEKA* a partir do algoritmo *id3*

A ferramenta *WEKA* construiu a árvore de decisão para o algoritmo *c4.5* como visualizado na Figura 19.

```
Classifier output
NmContribuintes = otimo: bem (12.0/2.0)
NmContribuintes = bom
| templo = sim
| | localização = boa: regular (3.0)
| | localização = ruim: bem (3.0/1.0)
| templo = nao: bem (6.0/1.0)
NmContribuintes = ruim
| localização = boa: mal (6.0)
| localização = ruim: regular (6.0/2.0)
NmContribuintes = pessimo: mal (12.0)

Number of Leaves :    7

Size of the tree :    11

Time taken to build model: 0.69 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      36          75    %
Incorrectly Classified Instances    12          25    %
```

Figura 19 – Árvore construída pela ferramenta *WEKA* a partir do algoritmo *c4.5*

7.4 IMPLEMENTAÇÃO

Para se implementar a ferramenta foram utilizados softwares de distribuição livre, *Personal Home Page* (PHP) como linguagem de programação sobre um servidor web Apache e MySQL versão 4.0.15 como sistema gerenciador de banco de dados (SGBD) referenciados no capítulo 3 deste trabalho. As páginas do sistema são estruturadas em *Hyper Text Markup Language* (HTML) e utiliza-se *JavaScript* (JS) para validar informações provindas dos formulários do lado cliente, conforme Goodman (2001).

7.4.1 TÉCNICAS E FERRAMENTAS UTILIZADAS

Para elaboração dos diagramas disponibilizados no trabalho foi utilizada a ferramenta *case Rational Rose 2002*.

Para o desenvolvimento dos formulários são utilizados os objetos disponíveis no próprio HTML: *input button* (botão), *input text* (caixa de texto de uma linha), *input hidden* (campo oculto), *select* (menu suspenso).

Foi utilizado também a ferramenta de *data mining Waikato Environment for Knowledge Analysis* (WEKA) para construção das árvores de decisão que foram aplicadas na mineração de dados do sistema conforme descrito no capítulo 7.3.

7.4.2 OPERACIONALIDADE DA IMPLEMENTAÇÃO

O sistema é baseado na web e após informado ao navegador o endereço na internet onde se encontra o mesmo é aberto a página inicial do sistema (Figura 20) requisitando o *login* e a senha do usuário para se obter o acesso ao sistema.



Figura 20 – Página inicial do sistema

Se a autenticação for bem sucedida, a sessão é inicializada e são apresentados os menus para navegação no sistema de acordo com o nível do usuário (Figura 21). As opções do menu são:

- a) cadastro: onde é apresentado os cadastro de dados necessários para o funcionamento do sistema;
- b) listagem: onde é apresentado a listagem dos dados já cadastrados no sistema;
- c) relatórios: onde é possível fazer refinamento nos dados cadastrados no sistema, escolhendo alguns atributos para efetuar a pesquisa;
- d) *data mining*: nesta opção é visualizada a página do *data mining* implementado no sistema;
- e) *logout*: opção para encerrar o sistema.



Figura 21 – Menus de navegação para o nível 1 (administrador global)

A diferença dos menus de navegação do nível 1 para o nível 2 são exclusivamente em relação à listagem e cadastro de usuários do sistema, ou seja, quando o nível do usuário for igual a 2 (dois) é omitido a opção usuário.

Se o nível do usuário for igual a 3 (três) o menu de navegação do sistema é apresentado de forma simplificada (Figura 22).

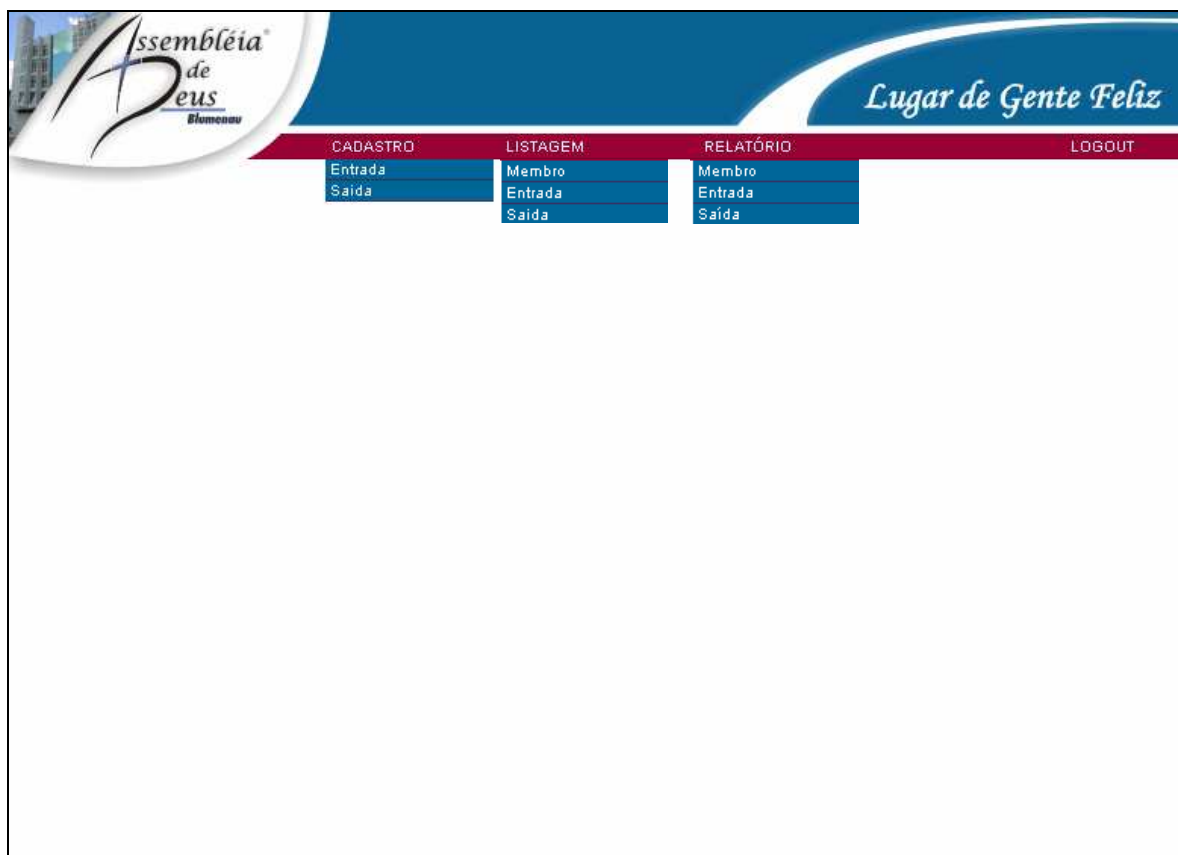


Figura 22 – Menus de navegação para o nível 3 (administrador congregacional)

Na Figura 23 é apresentado parcialmente a página de cadastro do membro da igreja, sendo que se o usuário for um administrador congregacional (nível 3), o campo congregação é omitido.

Os campos que possuem botões indicados por três pontos (...), possuem facilitadores de preenchimento, ou seja, ao clicar neste botão é aberta uma nova janela onde o usuário pode selecionar a opção desejada. Por exemplo, ao clicar no botão ao lado do campo naturalidade, é aberta uma nova janela com as cidades já cadastradas no sistema e após o usuário selecionar a cidade desejada o campo naturalidade é preenchido automaticamente (Figura 23). Estes recursos além de facilitar a utilização do sistema para o usuário garantem também a integridade referencial nos dados que serão gravadas no sistema.

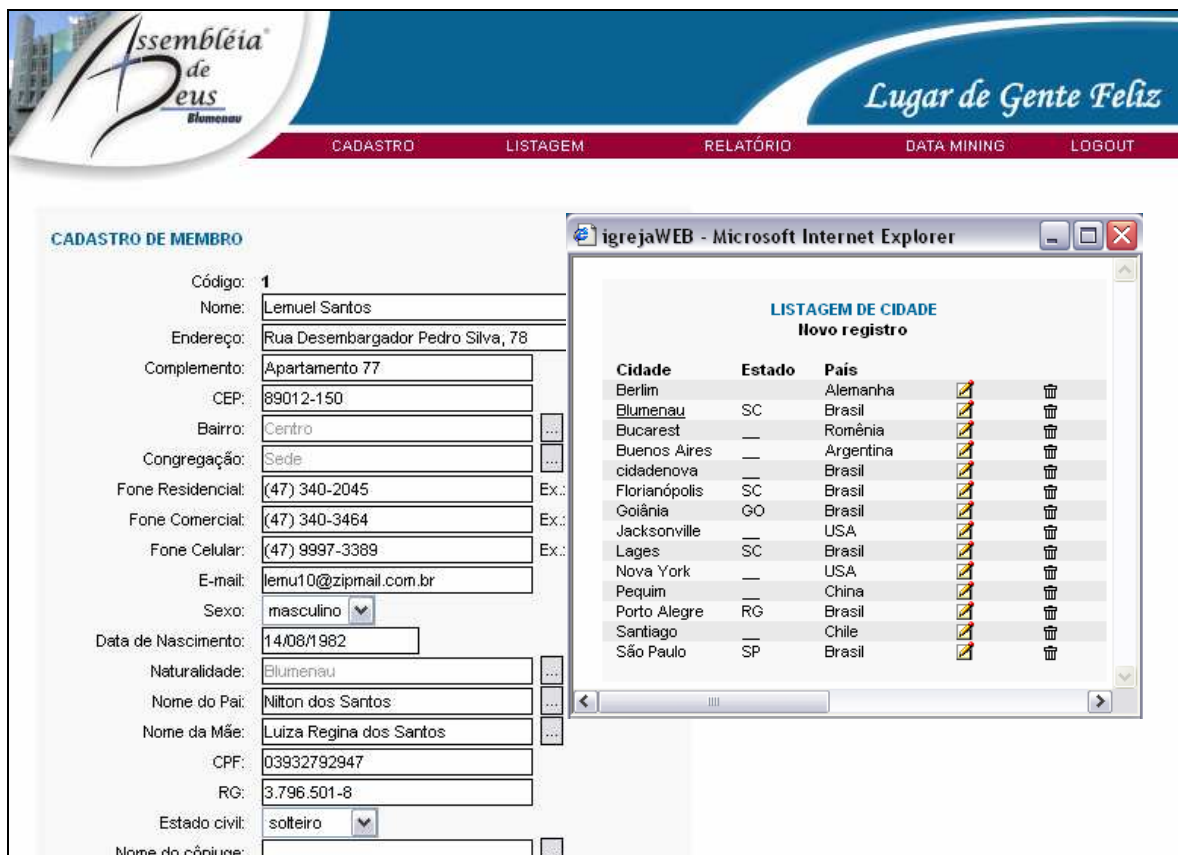


Figura 23 – Visualização parcial do cadastro de membro do sistema

O sistema possibilita a visualização dos dados do membro em uma nova janela (Figura 24). Este tipo de visualização é possível para as congregações e para os fornecedores.

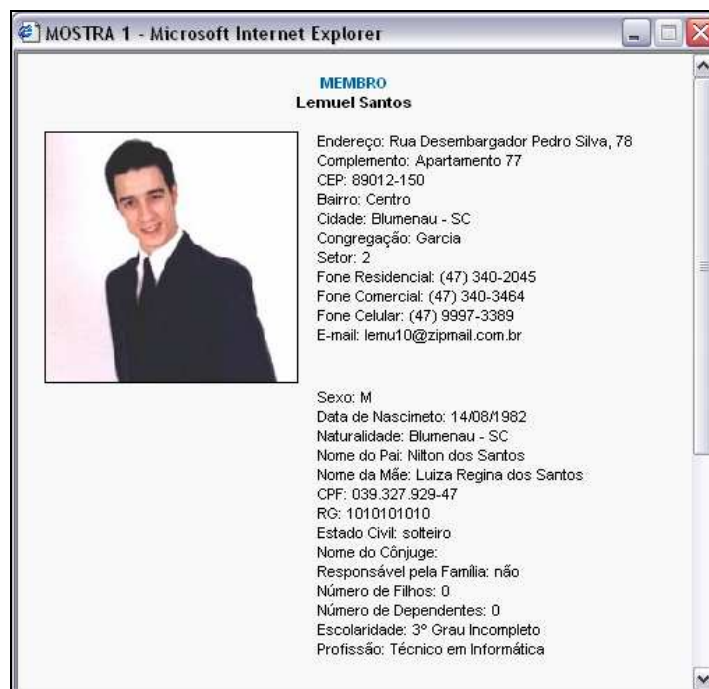


Figura 24 – Janela de visualização do membro da igreja

Na Figura 25 é apresentado simultaneamente o cadastro de entrada para os usuários de nível 1 e 2 (esquerda) em relação aos usuários de nível 3 (direita), sendo que para os usuários de nível 3, o campo congregação é omitido, pois ao efetuar o cadastro de alguma entrada a mesma será lançada automaticamente na congregação a qual este usuário pertence.

Figura 25 – Visualização do cadastro de entrada (nível 1 e 2 X nível 3)

Caso o usuário selecione a opção “imprimir recibo”, é gerada uma versão para impressão do recibo de entrada (Figura 26). Esta versão imprime na mesma página duas vias do recibo sendo que uma via deve ser entregue ao contribuinte e a outra é para uso da contabilidade da igreja.

Figura 26 – Visualização da versão para impressão do recibo de entrada

O mesmo processo no cadastro de entradas acontece também para o cadastro de saídas (Figura 27) onde se o nível do usuário for igual a 3 (três) é omitido o campo congregação.

Figura 27 – Visualização do cadastro de saída (nível 1 e 2 X nível 3)

Assim como no cadastro de entrada caso o usuário selecione a opção “imprimir recibo”, é gerada uma versão para impressão do recibo de saída (Figura 28). Esta versão imprime na mesma página duas vias do recibo de saída sendo que uma via deve ser entregue ao fornecedor e a outra é para comprovante da efetuação do pagamento ao fornecedor e para uso da contabilidade da igreja.

Figura 28 – Visualização da versão para impressão do recibo de saída

Na Figura 29 é apresentado um exemplo (listagem de cidade) de como são listados os dados cadastrados no sistema, sendo que para cada cadastro existe a sua respectiva listagem. Na listagem é possível editar, excluir os dados cadastrados, incluir um novo registro e ordenar a listagem pelos atributos exibidos no cabeçalho da listagem. Na Figura 29 a ordem padrão está por nome de cidade, porém pode se ordenar, por estado e também por país, sendo que este mecanismo funciona para todas as listagens do sistema.

Cidade	Estado	País		
Belo Horizonte	MG	Brasil		
Berlim		Alemanha		
Blumenau	SC	Brasil		
Brasília	DF	Brasil		
Campos	RJ	Brasil		
Curitiba	PR	Brasil		
Curitibaanos	SC	Brasil		
Florianópolis	SC	Brasil		
Fortaleza	CE	Brasil		
Fraiburgo	SC	Brasil		
Goiania	GO	Brasil		
Indaial	SC	Brasil		
Jacksonville		USA		
Lages	SC	Brasil		
Maceió	AL	Brasil		
Manaus	AM	Brasil		
Montes Claros	MG	Brasil		
Nova York		USA		
Pindamonhagaba	SP	Brasil		
Porto Alegre	RS	Brasil		
Recife	PE	Brasil		
Ribeirão Preto	SP	Brasil		
Rio de Janeiro	RJ	Brasil		
Salvador	BA	Brasil		
São Paulo	SP	Brasil		
Timbó	SC	Brasil		

Figura 29 – Visualização da listagem de cidade

Os relatórios servem para refinar os dados cadastrados no sistema para geração de uma listagem personalizada. O usuário coloca no campo os atributos necessários para a busca desejada e executa a pesquisa. Neste exemplo (Figura 30) foi executada uma pesquisa pelo nome da congregação. Após efetuada a pesquisa é possível ordenar os dados pelos atributos exibidos no cabeçalho da listagem, neste caso pode se ordenar por código do membro, nome do membro, tipo de membro, congregação e setor.

The screenshot shows the 'Assembleia de Deus' web application interface. At the top, there is a logo and navigation tabs: CADASTRO, LISTAGEM, RELATÓRIO, DATA MINING, and LOGOUT. The main content area contains a search form with fields for 'Código', 'Nome', 'Tipo de Membro', and 'Congregação'. Below the form are 'Pesquisar' and 'Listar todos' buttons. A section titled 'LISTAGEM DE MEMBRO' displays a table with the following data:

Código	Nome	Tipo de Membro	Congregação	Setor
8	Adão da Silva	Irmão	Garcia	2
1	Lemuel Santos	Presbítero	Garcia	2
4	Maria Albuquerque	Jovem	Garcia	2

Below the table, it indicates 'Quantidade: 3'.

Figura 30 – Visualização do relatório de membros

Na página da mineração de dados (Figura 31) do sistema deve-se escolher a congregação, o mês e o ano em que será efetuada a mineração de dados.

Após o preenchimento correto de todos os campos da página de mineração, deve-se escolher o algoritmo a ser utilizado na mineração, as opções são:

- algoritmo id3: como já apresentado no capítulo 7.3, apresenta maior processamento pois verifica todos os nós da árvore (sem podagem árvore), porém o resultado é mais preciso. Conforme Mena (1999, p.102) e conforme Amaral (2001,p.107), as regras podem ser geradas a partir da árvore de decisão (Quadro 8);

```

if ($algoritmo=="id3") {
  if ($nr_contribuintes=="otimo") {
    if ($nr_membros=="grande") {
      $sandamento="bem";
    }
    if ($nr_membros=="media") {
      $sandamento="bem";
    }
    if ($nr_membros=="pequena") {
      if ($templo=="sim") {
        $sandamento="regular";
      }
      else { //templo igual a não
        $sandamento="bem";
      }
    }
  }
}

```

```

}
} //fim contribuintes igual a ótimo
if ($nr_contribuintes=="bom") {
  if ($templo=="sim") {
    if ($localizacao=="boa") {
      $sandamento="regular";
    }
    if ($localizacao=="ruim") {
      if ($nr_membros=="grande") {
        $sandamento="bem";
      }
      if ($nr_membros=="media") {
        $sandamento="bem";
      }
      if ($nr_membros=="pequena") {
        $sandamento="regular";
      }
    }
  }
} //fim se templo igual a sim
if ($templo=="nao") {
  if ($nr_membros=="grande") {
    $sandamento="bem";
  }
  if ($nr_membros=="media") {
    $sandamento="bem";
  }
  if ($nr_membros=="pequena") {
    if ($localizacao=="boa") {
      $sandamento="regular";
    }
    if ($localizacao=="ruim") {
      $sandamento="bem";
    }
  }
} // fim templo igual a nao
} //fim nr_contribuintes igual a bom
if ($nr_contribuintes=="ruim") {
  if ($localizacao=="boa") {
    $sandamento="mal";
  }
  if ($localizacao=="ruim") {
    if ($nr_membros=="grande") {
      if ($templo=="sim") {
        $sandamento="regular";
      }
      if ($templo=="nao") {
        $sandamento="bem";
      }
    }
    if ($nr_membros=="media") {
      $sandamento="regular";
    }
    if ($nr_membros=="pequena") {
      if ($templo=="sim") {
        $sandamento="mal";
      }
      if ($templo=="nao") {
        $sandamento="regular";
      }
    }
  }
}
} // fim nr_contribuintes igual a ruim
if ($nr_contribuintes=="pessimo") {
  $sandamento = "mal";
}
} //fim se algoritmo = id3

```

Quadro 8 – Regras no sistema a partir da árvore de decisão gerada pelo algoritmo id3

- b) algoritmo c4.5: menor processamento, porém o resultado pode omitir algumas situações pois utiliza a técnica de podagem de árvore (Quadro 9).

```

// regras geradas a partir do algoritmo c45
if ($algoritmo=="c45") {
  if ($nr_contribuintes=="otimo") {
    $andamento="bem";
  }
  if ($nr_contribuintes=="bom") {
    if ($templo=="sim") {
      if ($localizacao=="boa") {
        $andamento="regular";
      }
      if ($localizacao=="ruim") {
        $andamento="bem";
      }
    }
    if ($templo=="nao") {
      $andamento="bem";
    }
  }
  if ($nr_contribuintes=="ruim") {
    if ($localizacao=="boa") {
      $andamento="mal";
    }
    if ($localizacao=="ruim") {
      $andamento="regular";
    }
  }
  if ($nr_contribuintes=="pessimo") {
    $andamento="mal";
  }
} //fim algoritmo c45

```

Quadro 9 – Regras geradas a partir da árvore de decisão gerada pelo algoritmo c4.5

Como exemplo foi efetuada uma mineração pelo algoritmo id3 (Figura 31) para a congregação Garcia durante o mês de setembro de 2004. Após a mineração é exibido o número de membros da congregação, a porcentagem de contribuintes em relação ao número de membros da congregação durante o mês selecionado, se a congregação possui um templo construído, a localização da congregação e o andamento da congregação (resultado da mineração).

Assembleia de Deus
Lugar de Gente Feliz

CADASTRO LISTAGEM RELATÓRIO DATA MINING LOGOUT

Congregação: Garcia ...

Mês: Setembro

Ano: 2004

Algoritmo: ID3 ou C4.5

RESULTADO DA MINERAÇÃO
Algoritmo id3

Número de Membros	Número de Contribuintes	Templo Construído	Localização	Andamento
3 = pequena	33.33% = ruim	sim	ruim	mal

Figura 31 – Visualização da página de mineração de dados do sistema

Com base nos resultados da mineração os administradores da igreja podem tomar decisões para melhorar o andamento da congregação, caso o resultado apresentado não seja satisfatório aos administradores da igreja, ou não estavam dentro das metas estabelecidas pela congregação.

7.5 RESULTADOS E DISCUSSÃO

O sistema foi disponibilizado na internet através do laboratório de informática do Campus IV da FURB para que os administradores da igreja e funcionários que trabalham atualmente na secretaria da IEADB tivessem acesso, utilizassem e testassem o sistema.

Após a apresentação do sistema aos administradores da igreja, os mesmos se mostraram muito otimistas em relação a idéia inovadora de possuir um sistema baseado na web, podendo descentralizar grande parte dos processos entre a congregação Sede e as demais congregações, possuindo a possibilidade de acompanhar o andamento de cada congregação mês a mês através da mineração dos dados.

Devido ao otimismo dos administradores da igreja em relação ao sistema, existe a real possibilidade de a partir de Janeiro de 2005, dependendo dos resultados de testes que serão efetuados em Novembro e Dezembro de 2004, o sistema implementado neste trabalho seja utilizado oficialmente pela IEADB.

Nos testes efetuados até o presente momento não foram relatados questões adversas (falhas, problemas, dificuldades e outros) no sistema, apenas sugestões para acrescentar determinadas praticidades (facilidades) aos usuários do sistema.

Confrontando o sistema proposto em relação aos sistemas já existentes no mercado apresentados no capítulo 6 deste trabalho, o sistema apresenta as seguintes vantagens:

- a) independência de plataforma: basta possuir um navegador para utilizar o sistema;
- b) instalação e atualização: nos outros sistemas (ROL-BL e GESIG) é necessário fazer a instalação e atualização de novas versões em cada máquina, já no sistema implementado neste trabalho só é necessária a instalação e atualização uma vez no servidor;
- c) descentralização da administração da igreja: com este sistema será possível aos administradores da congregação (nível 3) lançarem os dados da congregação no sistema e fazerem seus próprios relatórios, enquanto nos outros sistemas a administração é totalmente centralizada;
- d) facilidade de utilização: a utilização deste sistema é extremamente simples para os usuários que possuem um pré-conhecimento de como navegar em sites da internet, enquanto no ROL-BL possui uma interface gráfica com recursos gráficos do MS-DOS dificultando sua utilização;
- e) mineração de dados: no ROL-BL e no GESIG não possui nenhum tipo de mineração de dados, também não há registros que outros sistemas de administração de igrejas disponibilizem este recurso.

8 CONCLUSÕES

Uma igreja organizada deve possuir controle em relação aos membros da congregação, e aos processos administrativos que nela ocorrem. Estes controles organizacionais de uma igreja evangélica foram implementados no sistema apresentado neste trabalho, baseado nas peculiaridades da IEADB.

O sistema permitiu otimizar a forma de administração da IEADB, pois facilitará as transações de informações entre a congregação Sede e as demais congregações dos municípios, porém será necessário um breve treinamento aos administradores congregacionais para que estes façam bom uso do sistema.

Foram realizadas as validações e normalizações necessárias para que as informações cadastradas no banco de dados possuíssem integridade e consistências dos dados.

Fez-se uso da técnica de árvore de decisão para mineração de dados verificando o andamento da congregação, recurso este que pode servir de orientação para as tomadas de decisões dos administradores da igreja.

Avaliando as ferramentas utilizadas verificou-se que os softwares livres são de grande utilidade, pois retornam os resultados esperados e pré-estabelecidos de forma econômica e segura, o que torna uma grande vantagem para os desenvolvedores que estão iniciando seu caminho no mercado. As técnicas de DM implementadas no sistema se mostraram extremamente úteis na realidade de uma igreja, provando que é possível o uso de mineração de dados no mais diversos segmentos de negócios.

O sistema, mesmo na primeira versão, foi informalmente considerado pelos usuários melhor e mais prático em relação ao sistema utilizado atualmente na IEADB.

A conclusão deste trabalho e a implementação deste sistema representam à realização de um grande sonho de unir o aprendizado e a experiência adquirida durante os anos de academia e a realidade do dia a dia de uma igreja evangélica, e com muita satisfação ajudando a melhorar a forma de administração de uma igreja e abrindo caminho para o futuro da administração de igrejas evangélicas.

8.1 EXTENSÕES

A próxima tarefa é implementar no sistema o histórico do membro, sendo registrado no banco de dados todas as informações em relação as mudanças efetuadas no cadastro de cada membro e fazer a versão para impressão dos relatórios. Deve-se também implementar os *logs* do sistema, pois em um sistema baseado na web é essencial conhecer e possuir registrado a movimentação e as atividades de cada usuário que possui acesso ao sistema.

É necessário também implementar o controle de contas a pagar, pois algumas igrejas utilizam esse tipo de cadastro e disponibilizar uma outra forma para o preenchimento do endereço nos cadastros de membros, fornecedores e congregações a partir do CEP, como é feito em muitos sistemas e páginas da web.

Futuramente pode-se implementar no sistema o nível 4. Este nível seria o próprio membro da igreja, que poderia acessar o sistema para efetuar mudanças em seu cadastro, visualizar seu histórico e acessar informações sobre os eventos que acontecerão na igreja.

Os valores da mineração de dados que estão fixados no sistema poderiam se comportar de forma dinâmica sendo que o próprio usuário preencheria os valores para os atributos e informaria também os atributos que são relevantes ou não para a igreja em questão, fazendo desta maneira levaria o sistema a se comportar de acordo com a realidade de cada igreja.

Sugere-se para trabalhos futuros descobrir e implementar novas oportunidades para aplicação da mineração de dados em uma igreja, bem como utilização da filosofia de *data warehouse*.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, Fernanda Cristina Naliato do. **Data mining, técnicas e aplicações para o marketing direto**. São Paulo : Berkeley Brasil, 2001.
- ARBEX, Eduardo Compasso. **Iniciação científica ao data mining**. Resende, [2003]. Disponível em: < <http://www.inf.aedb.br/datamining/paginas/introducao.htm>>. Acesso em: 30 set. 2004.
- BERRY, Michael J. A. **Data mining techniques for marketing, sales and customer support**. New York: Wiley, 1997.
- BEZERRA, Eduardo. **Princípios de análise e projeto de sistemas com UML**. Rio de Janeiro: Campus, 2002.
- COMPOLT, Geandro Luis. **Sistemas de informação executiva baseado em data mining utilizando a técnica de árvores de decisão** 1999. 52 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) – Centro de Exatas e Naturais, Fundação Universidade Regional de Blumenau, Blumenau.
- CONVERSE, Tim. **PHP 4**, a bíblia. Tradução de Edson Furmankiewicz e Joana Figueiredo. Rio de Janeiro: Campus, 2001;
- CONALLEN, Jim. **Desenvolvendo aplicações web com UML**. Rio de Janeiro: Campus, 2003.
- DEITEL, H. M. **Internet e world wide web, como programar**. Tradução de Edson Furmankiewicz. Porto Alegre: Bookman, 2003.
- FAYYAD, Usama et al. **The KDD process for extracting useful knowledge from volumes of data**. Communications of the ACM, 1996.
- GOODMAN, Danny. **JavaScript**, a bíblia. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier, 2001.
- GWERCMAN, Sérgio. Evangélicos. **Super interessante**, São Paulo, n. 197, p. 52-61, fev. 2004.
- HARRISON, Thomas H. **Intranet data warehouse**. São Paulo: Berkely Brasil, 1998.
- HEUSER, Carlos Alberto. **Projeto de banco de dados**. Porto Alegre: Sagra Luzzatto, 2000.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. São Paulo, [2002]. Disponível em: <http://www.ibge.com.br/home/estatistica/populacao/censo2000/primeiros_resultados_amostra/brasil/pdf/tabela_1_1_2.pdf>. Acesso em: 08 ago. 2004.

LIMA, Adilson da Silva. **MySQL server, soluções para desenvolvedores e para administradores de banco de dados**. São Paulo: Érica, 2003.

LOH, Stanley. **Descoberta de conhecimentos em base de dados textuais**. Porto Alegre, [1997]. Disponível em: <<http://mozart.ulbra.tche.br/~loh/apostilas/dc-texto.htm>>. Acesso em: 20 set. 2004.

MENA, Jesus. **Data mining your website**. Woburn, MA: Butterworth-Heinemann, 1999.

MITCHELL T. **Machine Learning**. New York: McGraw-Hill, 1997.

NAVEGA, Sérgio. **Princípios essenciais do data mining**. São Paulo, [2002]. Disponível em: <<http://www.intelliwise.com/reports/i2002.pdf>>. Acesso em: 1 out. 2004.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**. São Paulo: Atlas, 1992

PRESSMAN, Roger S. **Engenharia de software**. Rio de Janeiro: McGraw-Hill, 2002.

QUINLAN, J. R. **Induction of decision trees**. Hingham, MA: Kluwer Academic Publishers, 1993.

SANTOS, Ismael dos. **Raízes da nossa fé**. Blumenau: Letra Viva, 1996.

SEGUNDO, Fábio R. **Programando para web com PHP**. Blumenau, [2004]. Disponível em: <<http://home.furb.br/fabio/tdsi-php-p1-v1d.pdf>>. Acesso em: 8 out. 2004.

SOARES, Wallace. **Crie um site B2C – business to consumer com PHP 4 e MySQL**. São Paulo: Editora Érica, 2001.

SOARES, Wallace. **Programando em PHP, conceitos e aplicações**. São Paulo: Érica, 2000.

STAIR, Ralph M. **Princípios de sistemas de informação**. Tradução de Maria Lúcia Lecker Vieira e Dalton Conde de Alencar; revisão técnica de Paulo Machado Cavalheiro e Cristina Bacellar. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A, 1998.

SUEHRING, Steve. **MySQL, a bíblia**. Tradução de Edson Furmankiewicz. Rio de Janeiro: Campus, 2002.

SZETO, Gong. **Interatividade na Web**. São Paulo: Berkeley Brasil, 1997.