

UNIVERSIDADE DO MINHO
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE SISTEMAS DE INFORMAÇÃO



***ZACCAR – SISTEMA DE CONHECIMENTO PARA APOIO À GESTÃO
DO RELACIONAMENTO COM CLIENTES***

Tese submetida à Universidade do Minho
para obtenção do grau de Doutor em Tecnologias e Sistemas de Informação, área de
especialização de Engenharia e Gestão de Sistemas de Informação,
elaborada sob a orientação do Professor Doutor João Álvaro Brandão Soares de Carvalho
e do Professor Doutor Manuel Filipe Vieira Torres dos Santos

MARIA DE FÁTIMA LEMOS FERREIRA ARMAS GONÇALVES

2003

Projecto parcialmente financiado por uma bolsa do PRODEP II,
medida 5, acção 5.2, concurso nº1/96, Doutoramentos.

À memória do meu pai

Agradecimentos

Para a elaboração deste trabalho, beneficiei de muitos contributos das mais variadas formas, quer na troca de opiniões, no apoio logístico ou, simplesmente, no apoio amigo. Na impossibilidade de agradecer a todos, referencio aqueles que, de algum modo, mais se destacaram.

Em primeiro lugar, um agradecimento especial ao Professor Doutor João Álvaro Carvalho, meu orientador desde o início, pelos conhecimentos que me transmitiu e pelos seus conselhos e sugestões sempre oportunos. Igualmente, ao Professor Doutor Manuel Filipe Santos cuja orientação, embora tenha começado apenas a meio do percurso, não deixou de ser igualmente válida e empenhada. A ambos o meu agradecimento.

Ao Departamento de Sistemas de Informação da Universidade do Minho por ter tornado possível este projecto e pela simpatia que, em todas as ocasiões, muitos dos seus elementos me manifestaram.

Aos meus colegas de Doutoramento e, em particular, à Paula Morais e à Filomena Lopes, pelo entusiasmo que me transmitiram, pela sua disponibilidade e amizade.

Ao Instituto Superior de Contabilidade e Administração de Coimbra por me ter proporcionado as facilidades necessárias para a execução desta tese.

Um agradecimento muito especial à Rosália que me incentivou a iniciar este projecto e me acompanhou nos primeiros passos necessários à sua concretização.

Aos meus colegas do grupo de Informática, Paulo Belfo pela cedência de dados para que me fosse possível treinar o *Clementine* e ao André Rodrigues pela sua permanente

disponibilidade para a instalação e manutenção dos sistemas informáticos de que necessitei e pelas suas sugestões em vários pontos da tese.

À Estela Trigo e ao Luís Amaral pela colaboração nas formulações matemáticas.

Aos responsáveis da AMMA por terem posto à minha disposição, incondicionalmente, os dados da sua empresa, contribuindo, assim, para que eu pudesse provar a viabilidade da ferramenta desenvolvida.

À minha família por toda a atenção que me mereciam e não lhes dispensei.

Um agradecimento especial à minha filha Micaela que, com as suas sugestões e participação, tornou possível o aspecto gráfico das figuras que se encontram ao longo da tese.

Aos meus pais, pela forma com que me criaram e a educação que me deram, permitindo que me tornasse na pessoa que hoje sou.

Fátima Gonçalves

2001

Resumo

Nesta tese apresenta-se uma ferramenta para gestão, integração e consolidação do Conhecimento sobre o Comportamento dos Clientes (CCC) obtido a partir da actuação de ferramentas de *data mining* sobre bases de dados transaccionais de organizações.

As ferramentas de *data mining* permitem automatizar a detecção de padrões de comportamento dos clientes de uma organização a partir das bases de dados transaccionais, num processo designado por Descoberta do Conhecimento em Bases de Dados (DCBD).

Estes padrões podem ser transmitidos aos agentes organizacionais e utilizados em campanhas de *marketing* e outras actividades no contexto da organização.

No entanto, este conhecimento sobre o comportamento dos clientes não é, normalmente, objecto de qualquer tratamento que permita a análise das razões para o seu aparecimento ou da sua evolução bem como a consolidação com outro conhecimento sobre o CCC já existente.

Há, pois, neste processo, uma situação que consideramos que pode ser melhorada com a introdução dum novo conceito - a **Gestão do CCC** – o qual conduz a uma nova actividade organizacional – **Zelar pelo CCC**.

A Gestão do Conhecimento sobre o Comportamento dos Clientes é entendida como a confrontação deste conhecimento com outro conhecimento já existente na organização, resolvendo potenciais conflitos, actualizando-o e acrescentando explicações pertinentes para a evolução temporal verificada.

As principais contribuições deste trabalho centram-se:

- na apresentação do conceito “a Gestão do Conhecimento sobre o Comportamento dos Clientes” que conduz a uma nova tarefa organizacional “Zelar pelo Conhecimento sobre o Comportamento dos Clientes”;

- no estabelecimento de uma forma de estrutura do CCC e seu registo;
- na concepção e exploração dum sistema de conhecimento para apoio à gestão do conhecimento do comportamento dos clientes - o sistema **ZACCAR** (**Z**elar pela **A**quisição do **C**onhecimento dos **C**lientes, sua **A**ctualização e **R**egisto) - cujo objectivo principal é permitir a viabilidade da nova tarefa através da:
 - recolha e uniformização dos padrões de comportamento obtidos com uma ferramenta de *data mining*;
 - confrontação desses padrões com o conhecimento já existente acerca do comportamento dos clientes, actualizando-o;
 - validação e documentação, pelo gestor do conhecimento organizacional, do conhecimento já actualizado;
 - integração do conhecimento depois de actualizado e completado, numa base de conhecimento que fará parte integrante do conhecimento organizacional;
- no processo de consolidação do conhecimento descoberto em bases de dados, resolvendo problemas de interpretação, integração e conflitos.

Na prossecução dos objectivos que estiveram presentes na elaboração deste trabalho, foi feita uma análise pormenorizada da prática de CRM (*Customer Relationship Management*) e sua relação com o conhecimento organizacional bem como do CCC com ênfase no tratamento que é dado a este conhecimento.

O sistema ZACCAR pode-se considerar um sistema inovador uma vez que permite às organizações dispor de uma base de conhecimento do CCC, actualizada dum forma semi-automática onde está, ainda, registada uma evolução dos padrões de comportamento dos clientes e que faz parte integrante do conhecimento organizacional.

Um protótipo do ZACCAR foi desenvolvido, recorrendo a tecnologia existente; para demonstrar a sua exequibilidade, foram efectuados dois estudos de casos os quais demonstram que o sistema possui potencialidades interessantes que se poderão tornar muito úteis em qualquer empresa onde o sistema seja implantado quer como sistema independente quer como integrado noutros sistemas empresariais de maior abrangência.

Palavras chave: Bases de Conhecimento, *Business Intelligence*, Conhecimento Organizacional, Conhecimento sobre o Comportamento dos Clientes, *CRM*, *Data Mining*, Descoberta do Conhecimento em Bases de Dados.

Abstract

In this thesis it is presented a tool to take care of the Customers' Behaviour Knowledge (CBK) obtained when a data mining tool acts in the organisational databases to manage and integrate it in the organisational knowledge, through a consolidation process with the existing knowledge.

Data mining tools automate the detection of customers'behaviour patterns from the organisational databases in a process called Knowledge Discovery in Databases (KDD). These patterns may be transmitted to organisational agents and used in marketing campaigns and other activities in the organisation.

However, the CBK is not usually treated to allow the analysis why it exists or how it evolves as well its consolidation with other existing CBK.

So, we consider that, in this process, there is a situation that can be optimized through the introducing of a new concept - the management of CBK - conducting to a new organisational activity - to take care of the CBK.

The management of CBK is intended as the confrontation of this knowledge with other existing knowledge, resolving potential conflicts, updating it and adding pertinent explanations to the temporal evolution of the customers'behaviour patterns.

The most important contributions of this work are:

- the presentation of the concept "The management of the Customers' Behaviour Knowledge" that allows a new organisational task: "To take care of the Customers' Behaviour Knowledge";

- the creation and exploration of a knowledge system to help the management of the CBK - the ZACCAR system - whose main objective is to permit the viability of the new task that is got by:
 - the collecting and uniformization of the behaviour patterns obtained with a data mining tool;
 - the confrontation of these patterns with existing CBK, updating it;
 - the validation and documentation, by the manager of the organisational knowledge, of the knowledge after to be updated;
 - the integration of the updated knowledge in a knowledge base that will be an integrant part of the organisational knowledge;
- in the consolidation process of the knowledge discovered in databases, resolving interpreting and integration problems as possible conflicts.

Attending the objectives considered in this work, it was made a detailed analysis of the practice of CRM (Customer Relationship Management) and its relation with the organisational knowledge as well of the CBK with emphasis in the treatment given to this knowledge

ZACCAR can be considered an innovating system as, with it, the organisations can have a knowledge base of the CBK, updated in a semi-automatic process where it can be yet, stored the evolution of the customers'behaviour patterns and turned as an integrant part of the organisational knowledge.

It was developed a prototype of ZACCAR, using existing technology; to prove its feasibility it was conducted two case studies; these cases showed that the system has good potentialities that will be very useful in an enterprise where the system can be implemented either as independent system or integrated in other organisational systems with a greater covering.

Key words: Business Intelligence, CRM, Customers' Behaviour Knowledge, Data Mining, Knowledge Base, Knowledge Discovery from Databases, Organisational Knowledge.

Índice

Índice de Figuras	v
Índice de Tabelas	vii
Notação	viii
Acrónimos	viii
Símbolos Gerais e Abreviaturas	ix
1. Introdução	1
1.1 Motivação e Objectivos	3
1.2 Contribuições	8
1.3 Processo de Investigação	11
1.4 Estrutura da Tese	13
2. CRM e Conhecimento do Comportamento dos Clientes.....	16
2.1 A Prática de CRM.....	17
2.2 O Cliente	26
2.3 O Conhecimento sobre o Comportamento dos Clientes (CCC)	30
2.4 O CCC nas Empresas.....	32
2.5 Razões para o Aumento da Importância do CCC.....	37
2.6 Obtenção do CCC	39
2.7 CCC e Conhecimento Organizacional.....	42
2.8 O Percurso do CCC	44
2.9 Sumário.....	47
3. Tecnologias para Aquisição e Registo de Conhecimento.....	49
3.1 Detecção de Eventos e Criação de Registos	50
3.2 Tecnologias para Registo de Factos	53
3.2.1 As Bases de Dados.....	53
3.2.2. O <i>Data Warehouse</i>	60
3.3 Criação e Representação do Conhecimento Conceptual	69
3.3.1. A Descoberta de Conhecimento em Bases de Dados	70
3.3.2. As Ferramentas de <i>Data Mining</i>	77

3.3.3. Modelos para Representação do Conhecimento Obtido com <i>Data Mining</i>	82
3.4 O Armazenamento do Conhecimento Conceptual	90
3.5 Sumário	91
4. Gestão do Conhecimento do Comportamento dos Clientes	92
4.1 Modelo Global do Processo de Aquisição, Registo e Exploração do CCC	94
4.2 O Cenário do Tratamento dos Padrões	96
4.3 Suporte Tecnológico para Zelar pelo CCC	98
4.3.1 A Base de Conhecimento do Comportamento dos Clientes	101
4.3.2 O Gestor de Conhecimento do Comportamento dos Clientes	104
4.4 A Viabilidade da Proposta	108
4.5 Sumário	108
5. O Sistema ZACCAR	110
5.1 Opções tecnológicas	110
5.1.1 Representação do Conhecimento através de Regras de Produção	111
5.1.2 A linguagem de programação em lógica <i>Prolog</i>	116
5.2 A Arquitectura Funcional do Sistema	119
5.2.1 O Gestor de Conhecimento do Comportamento dos Clientes	120
5.2.2 A Base de Conhecimento do Comportamento dos Clientes	134
5.2.3 Consulta à BCCC	134
5.3 Sumário	136
6. Viabilidade do Sistema ZACCAR	137
6.1 Caso 1 - Turismo	139
6.1.1. Momento 1	142
6.1.2. Momento 2	144
6.1.3. Momento 3	147
6.1.4. Momento 4	149
6.1.5. Padrões de comportamento correspondentes a reacções negativas às campanhas	152
6.1.6. Discussão	153
6.2 Caso 2 - Indústria de Confecção	155
6.2.1 Fase 1	156
6.2.2 Fase 2 - Estudo ao Longo de Três Momentos	162
6.2.3 Discussão	174
6.3 Situações não Previstas Detectadas no Estudo de Casos	175
6.4 Sumário	178
7. Conclusões	180
7.1 Síntese do Trabalho Desenvolvido	180
7.2 Perspectivas futuras	184
Referências Bibliográficas	187

Anexos.....	195
Anexo 1 - Símbolos utilizados nas figuras	197
Anexo 2 – Créditos	203
Anexo 3 – Alguns módulos dos programas desenvolvidos.....	205
Anexo 3.1 - Programa <i>confronto</i>	207
Anexo 3.2 - Programa <i>interface</i>	209
Anexo 3.3 - Programa <i>assimilação</i>	212
Anexo 4 – Inquéritos para obtenção dos dados utilizados no caso 1	215
Anexo 5 – Extracto da base de dados utilizada no caso 1	221
Anexo 6 – Regras obtidas, no caso 1, com o <i>Clementine</i>	225
Anexo 6.1 Regras obtidas no momento 1	227
Anexo 6.2 Regras obtidas no momento 2.....	228
Anexo 6.3 Regras obtidas no momento 3	229
Anexo 6.4 Regras obtidas no momento 4.....	230
Anexo 7 – Extracto da base de dados usada no caso 2.....	231
Índice de Autores	235
Índice Remissivo	237

Índice de Figuras

Figura 1.1- Relacionamento desejável entre o conhecimento organizacional onde está incluído o CCC, CRM e BI.	4
Figura 1.2 – Arquitectura dum sistema de conhecimento (Adaptado de [Stefix, 1995]).	6
Figura 1.3 - Estrutura da tese.....	13
Figura 2.1 – Pirâmide de clientes reais e potenciais (Adaptado de [Curry & Curry, 2000]).	20
Figura 2.2 – Ciclo de vida cliente-empresa (Adaptado de [Brown, 2000]).	28
Figura 2.3 - Questões fundamentais para o conhecimento do comportamento dos clientes [Berson <i>et al.</i> , 1999].	31
Figura 2.4 - Percurso do CCC.	45
Figura 3.1 - Tecnologias para obtenção de eventos e sua transformação em conhecimento.....	49
Figura 3.2 - O SGBD.....	54
Figura 3.3 – Uma tabela de bases de dados.....	56
Figura 3.4 – <i>Array</i> tridimensional.	60
Figura 3.5 – Integração de dados num <i>data warehouse</i> (Adaptado de [Inmon & Hackathorn, 1994]).	63
Figura 3.6 – Organização dos dados num <i>data warehouse</i> (Adaptado de [Inmon & Hackathorn, 1994]).	65
Figura 3.7 – O processo de DCBD.....	72
Figura 3.8 - O processo de DCBD usando o conhecimento organizacional.	74
Figura 3.9 - A integração do <i>data warehouse</i> no processo de DCBD.....	76
Figura 3.10 – Exemplo de uma árvore de decisão.....	83
Figura 4.1 - Modelo global do processo de aquisição, registo e exploração do CCC.....	96
Figura 4.2 Modelo global do processo de aquisição, actualização, memorização e exploração do CCC.....	99
Figura 4.3 – O sistema ZACCAR.....	99
Figura 4.4 - ZACCAR - Uma ponte entre DM e o conhecimento organizacional.	100
Figura 5.1 – Arquitectura do sistema de conhecimento ZACCAR.	120
Figura 5.2 - O processo do global do GCCC e respectivas fases	121
Figura 5.3 - Procedimento para gerar o compilador.....	123
Figura 5.4 - Procedimento "Uniformização de formatos".....	124
Figura 5.5 – Procedimento “Confronto”.	125
Figura 5.6 - Procedimento “ <i>Interface</i> ”.....	128
Figura 5.7 - Opções a serem seleccionadas pelo gestor do conhecimento.	129
Figura 5.8 - Os procedimentos "Pesquisa" e "Assimilação"	130
Figura 5.9 – Os processos envolvidos no GCCC.	133
Figura 5.10 – A consulta à BCCC.	135
Figura 6.1 – <i>Stream</i> implementada no <i>Clementine</i> no primeiro caso.	140

Figura 6.2 – Metodologia seguida no estudo de caso.....	156
Figura 6.3 – Código do produto.	160
Figura 6.4 – <i>Stream</i> implementada no <i>Clementine</i> no segundo caso.	163
Figura 6.5 – Alteração a introduzir no procedimento "Assimilação".....	178

Índice de Tabelas

Tabela 2.1 - Eventos na vida dum cliente e iniciativas de CRM que podem desencadear.	24
Tabela 3.1 - Tabela de clientes.	57
Tabela 3.2 - Tabela de produtos.	57
Tabela 3.3 - Tabela de transacções.	57
Tabela 3.4 - Agregação de clientes com transacções.	65
Tabela 3.5 - Agregação, por mês, de produtos vendidos por código postal.	66
Tabela 3.6 - Agregação de produtos vendidos por profissão.	66
Tabela 3.7 - Agregação de produtos vendidos por localidade.	66
Tabela 5.1 - Tecnologias utilizadas no desenvolvimento do protótipo.	111
Tabela 5.2 - Simbologia EBNF utilizada na sintaxe das regras.	113
Tabela 5.3 - Tarefas do procedimento “Interface”.	128
Tabela 5.4 - Tarefas no procedimento “Pesquisa”.	131
Tabela 5.5 - Correspondência entre os conjuntos de regras e os ficheiros que os materializam.	133
Tabela 6.1 - Um subconjunto dos dados das campanhas de <i>marketing</i>	140
Tabela 6.2 - Atributos da base de dados.	140
Tabela 6.3 - A distribuição de cada atributo no momento 1.	142
Tabela 6.4 - A distribuição de cada atributo no momento 2.	145
Tabela 6.5 - A distribuição de cada atributo no momento 3.	147
Tabela 6.6 - A distribuição de cada atributo no momento 4.	149
Tabela 6.7 - Estrutura da tabela de clientes.	159
Tabela 6.8 - Estrutura da tabela de artigos.	159
Tabela 6.9 - Estrutura da tabela de cabeçalho de encomendas.	159
Tabela 6.10 - Estrutura da tabela de pedidos.	159
Tabela 6.11 - Valores de campos no código dum produto.	161

Notação

As expressões e palavras em itálico correspondem a termo ingleses que são utilizados na linguagem corrente.

Como se pode observar na lista seguinte, há dois tipos de acrónimos que foram inseridos no texto, uma primeira vez, de maneira diferente:

- Os que fazem parte da linguagem corrente e são sobejamente conhecidos (ATM, BI, EBNF, EDI, CRM, GRI, OLAP, SQL) são escritos, seguidos da respectiva descrição entre parêntesis.

Exemplo: CRM (*Customer Relationship Management*)

- Os que definimos neste trabalho (BCCC, CCC, DCBD, GCCC, ZACCAR) ou os que nem sempre são usados (DM, HCI, SGBD) para os quais é escrita a respectiva descrição, acompanhada do acrónimo entre parêntesis.

Exemplo: Conhecimento do Comportamento dos Clientes (CCC).

Acrónimos

ATM	<i>Automatic Transaction Machines</i>
BCCC	Base de Conhecimento do Comportamento dos Clientes
BI	<i>Business Intelligence</i>
CCC	Conhecimento do Comportamento dos Clientes
CRM	<i>Customer Relationship Management</i>
DCBD	Descoberta de Conhecimento em Bases de Dados
DM	<i>Data Mining</i>

EBNF	<i>Extended Backus-Naur Form</i>
EDI	<i>Electronic Data Interchange</i>
GCCC	Gestor de Conhecimento do Comportamento dos Clientes
GRI	<i>Generalised Rule Induction</i>
HCI	<i>Human Computer Interface</i>
OLAP	<i>On-Line Analytical Processing</i>
SGBD	Sistema de Gestão de Bases de Dados
SQL	<i>Structured Query Language</i>
ZACCAR	Zelar pela Aquisição do Conhecimento dos Clientes, sua Actualização e Registo

Símbolos Gerais e Abreviaturas

E	Expressão
L	Linguagem
X, Y, Z	Atributos de uma relação
\mathcal{P}	Ocorrência de um padrão
\mathcal{R}	Domínio de um padrão
R	Conjunto de atributos
A_i	Atributo de uma relação
Dom_{A_i}	Domínio do atributo A_i
a_i	Uma ocorrência de Dom_{A_i}
α	Função definida sobre Dom_{A_i}
β	Função definida sobre um produto cartesiano de domínios
p	Coluna de uma tabela de uma base de dados
x	Linha de uma tabela de uma base de dados

$p(x)$	Ocorrência na coluna p e linha x de uma tabela de uma base de dados
e	Evento
$e.T$	Tempo do evento e
$e.A$	Um atributo A do evento e
U	Sequência de eventos
Q_i	Episódio i
x_i	Variável de um evento
S	Conjunto alvo
C_i	Classe de dados semelhantes (subconjunto de S)
D_i	Descrição da classe C_i
$cond_i$	Condição da classe C_i
r	Regra
r_i	Regra i
n_1	Percentagem dos elementos de S que verificam os antecedentes da regra r .
n_2	Percentagem de n_1 que também verifica o conseqüente da regra r .
descriptorA	Antecedente de uma regra
descriptorB	Conseqüente de uma regra
extensão	Parte estendida de uma regra
data_i	Data de actualização de uma regra canónica estendida
comentario_i	Comentário do gestor do conhecimento a inserir na regra r na data _i
parâmetros_i	Grau de incerteza da regra r na data _i
suporte_i	Nível de suporte da regra r na data _i
confiança_i	Nível de confiança da regra r na data _i
temp_i	Ficheiro temporário para registo de regras

<i>padrões_dm</i>	Ficheiro de registo das regras obtidas com uma ferramenta de <i>data mining</i> .
<i>F</i>	Conjunto cujos elementos são regras
<i>F_i</i>	Conjunto cujos elementos são regras
<i>F_i^j</i>	Subconjunto do conjunto <i>F_i</i>
<i>unif</i>	Função <i>uniformização</i> de regras que transforma as regras obtidas com <i>data mining</i> em regras canónicas
<i>conf₁</i>	Uma das funções do procedimento <i>confronto</i> de regras que confronta as novas regras com regras já existentes actualizando-as.
<i>conf₂</i>	Uma das funções do procedimento <i>confronto</i> de regras que transforma uma nova regra em regra canónica estendida.
<i>interface</i>	Função que assegura o <i>interface</i> entre uma regra e o gestor do conhecimento
<i>pesquisa</i>	Função <i>pesquisa</i> de regras semelhantes
<i>k_i</i>	Comentário do gestor do conhecimento

1.Introdução

"The fundamental problem of understanding intelligence is not the identification of a few powerful techniques, but rather the question of how to represent large amounts of knowledge in a fashion that permits their effective use and interaction"
(Goldstein and Papert)¹

A gestão é uma ciência de que decorre um conjunto de técnicas que procuram articular recursos disponíveis para atingir determinados objectivos [Zorrinho, 1991].

Uma organização² obtém e regista em bases de dados os dados obtidos nos vários pontos de contacto com os seus clientes; estes registos podem fornecer, à organização, Conhecimento sobre o Comportamento dos seus Clientes (CCC) o qual levará a um eficaz processo da **Gestão do Relacionamento com os Clientes**.

Ferramentas de *Data Mining*³ (DM) [Agrawal *et al.*, 1993], [Fayyad *et al.*, 1996a], [Han & Kamber, 2001] actuando sobre as bases de dados transaccionais das organizações conduzem à obtenção de padrões⁴ de comportamento dos clientes os quais

¹ apud [Stefix, 1995].

² Uma *organização* é uma unidade social deliberadamente construída para alcançar fins específicos num dado contexto social [Etzioni, 1980]. Sendo uma empresa, enquanto unidade social organizada, uma organização e tendo em atenção o âmbito deste trabalho, usaremos, indiferentemente, os termos organização e empresa para designar o mesmo tipo de entidade.

³ *Data mining* é definido como a tarefa de analisar e explorar grandes quantidades de dados com vista à obtenção de conhecimento previamente desconhecido.

⁴ Um *padrão* é uma expressão *E*, numa linguagem *L*, descrevendo um conjunto de factos, i.e., ocorrências numa base de dados [Fayyad *et al.*, 1996].

tornando-se parte do conhecimento organizacional⁵, têm em vista contribuir para que as organizações atinjam os objectivos pretendidos.

Aquelas ferramentas conduzem a resultados que podem ser apresentados em diferentes formatos, frequentemente de difícil interpretação.

Os gestores não possuem, regra geral, conhecimentos técnicos que lhes permitam interpretar os modelos de DM nem tão pouco explorar os resultados na forma em que são obtidos. Por este motivo, muito do conhecimento obtido com DM ou nunca é utilizado ou sendo-o não o é convenientemente.

Gera-se, assim, uma lacuna entre o processo de aquisição do conhecimento obtido através de ferramentas de DM e a sua comunicação aos gestores do negócio a fim de que o possam utilizar em decisões conducentes a uma melhoria do seu negócio.

O processo aqui apresentado e desenvolvido tem como objectivo estabelecer a ponte necessária entre DM e o conhecimento organizacional, permitindo, por um lado, a uniformização dos resultados de DM e, por outro lado, a sua integração, duma forma semi-automática, no conhecimento organizacional, criando, ainda, condições que potenciem a sua consulta duma forma eficiente ou a sua integração automática noutras aplicações da organização.

Assim, o principal contributo deste trabalho é, considerando o interesse para as organizações do conceito a "**Gestão do CCC**", uma proposta de uma nova actividade organizacional "**Zelar pelo CCC**" que, para ser viável, carece duma ferramenta que caracterizamos e para a qual desenvolvemos um protótipo.

Tendo em atenção que *Business Intelligence* (BI) - é um conjunto de conceitos, métodos e processos que permitem melhorar as decisões do negócio, usando informação proveniente de múltiplas fontes e aplicando a experiência e formulação de hipóteses

⁵ "**Conhecimento organizacional** é algo de que os agentes organizacionais poderão precisar de saber para executar as suas acções." [Carvalho & Morais, 2001].

para desenvolver uma compreensão correcta da dinâmica do negócio [Brackett, 1999], esta tese representa um contributo para a BI, inserindo-se, claramente, nesta área.

1.1 Motivação e Objectivos

O termo gestão pode ser utilizado em perspectivas e contextos variados; na elaboração deste trabalho assumem particular relevância: a Gestão do Conhecimento do Comportamento dos Clientes e a Gestão do Relacionamento com os Clientes.

Entendemos a *Gestão do CCC* como a confrontação deste conhecimento com anterior conhecimento do mesmo tipo já existente na organização, resolvendo potenciais conflitos, actualizando-o e acrescentando justificações pertinentes para a evolução temporal verificada.

A *Gestão do Relacionamento com os Clientes*, usualmente conhecida como CRM (*Customer Relationship Management*), é um processo que, baseado na evolução do comportamento do cliente e na aprendizagem obtida a partir de cada contacto (e.g., compras, consultas, reclamações), tem em vista um tratamento personalizado do cliente e o fortalecimento das relações cliente-empresa.

Este processo assume uma maior importância quando se passa duma estratégia de negócio centrada no produto para uma estratégia de negócio centrada no cliente [Newell, 2000a].

O maior proveito da CRM é, sem dúvida, a fidelização dos clientes da empresa e a aquisição de potenciais clientes com perfis semelhantes àqueles.

Para que a CRM tenha sucesso no contexto das organizações, torna-se indispensável que estas possuam um conhecimento actualizado sobre o comportamento dos seus clientes (quer reais quer potenciais) e sobre a sua evolução ao longo do tempo.

Uma boa prática de CRM, a par dum conhecimento organizacional dinamicamente actualizado do qual faz parte integrante o CCC, constitui um bom contributo para a BI (Figura 1.1) [Dhar & Stein, 1997], [Kudyba & Hoptroff, 2001].

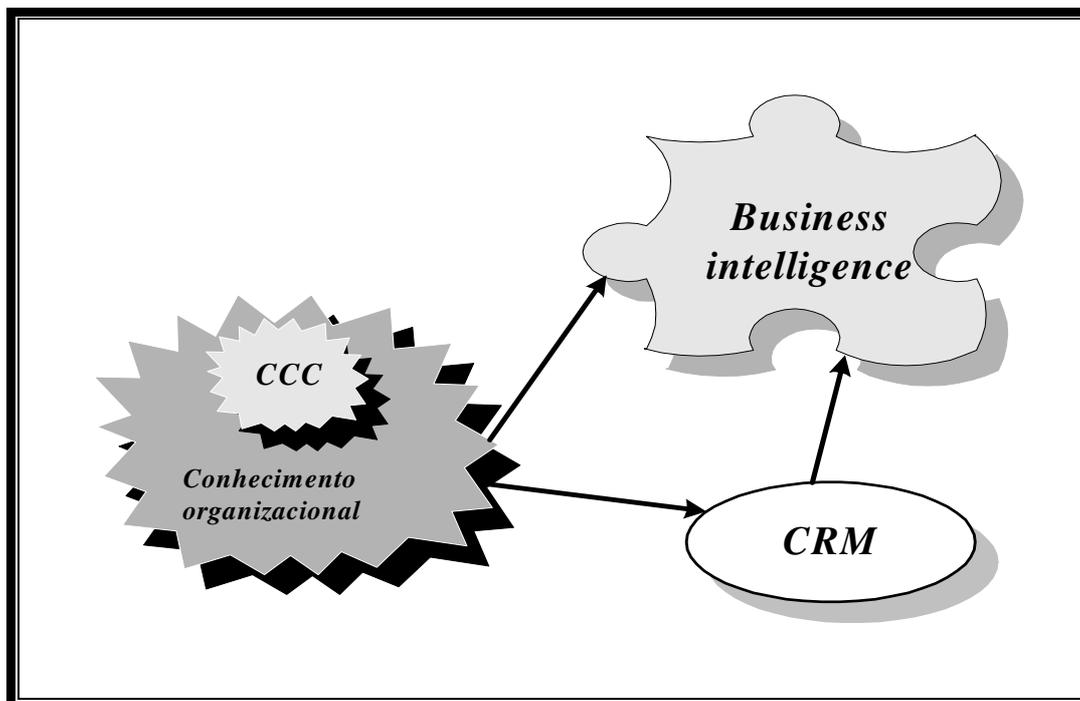


Figura 1.1- Relacionamento desejável entre o conhecimento organizacional onde está incluído o CCC, CRM e BI.

Tomando como base uma revisão da literatura sobre CRM ([Berson *et al.*, 1999], [Brown, 2000], [Newell, 2000a], [Berry & Linoff, 2000], [Burnett, 2001], www.crm_forum.com, entre outros), constatámos que:

- as empresas necessitam, cada vez mais, de adoptar uma prática de CRM pois a capacidade de atrair e conservar os clientes é, no actual mundo dos negócios, fundamental para a sua sobrevivência;
- a adopção de tal prática passa pela obtenção e gestão do CCC;
- esse conhecimento é, cada vez mais, obtido, duma forma automática e semi-automática, pela actuação de ferramentas de DM em processos de **D**escoberta de **C**onhecimento em **B**ases de **D**ados (DCBD) [Piatetsky-Shapiro & Frawley, 1991], [Frawley *et al.*, 1991], [Brachman & Anand, 1996], [Fayyad *et al.*, 1996d] onde se

encontram armazenados os registos das interacções entre a empresa e os seus clientes;

- o conhecimento assim obtido não é, normalmente, gerido, tendo em atenção zelar pelo novo conhecimento bem como pela actualização do já existente, permitindo, a qualquer momento, o acesso não só a uma regra de comportamento dos clientes mas também à sua evolução ao longo do tempo.

Tendo constatado, pelas leituras efectuadas e casos reais observados, a ausência duma gestão deste conhecimento tal como entendemos que ela deva ser, tentamos mostrar o interesse, para a organização, do conceito a “**Gestão do CCC**” o qual conduz a uma nova actividade organizacional “**Zelar pelo CCC**”.

Com o objectivo de viabilizar esta nova actividade organizacional, desenvolvemos um sistema de conhecimento que permite estabelecer a ponte entre os padrões obtidos pelas ferramentas de DM e o conhecimento organizacional onde o CCC deve estar integrado. Ao sistema desenvolvido chamamos ZACCAR (**Z**elar pela **A**quisição do **C**onhecimento dos **C**lientes sua **A**ctualização e **R**egisto).

Um **sistema de conhecimento** [Stefix, 1995] é tipicamente composto por um conjunto de recursos - *hardware*, *software* e humanos – aos quais compete a criação e manutenção de uma base de conhecimento, garantindo a actualização, segurança e integridade desta e disponibilizando-a para que possa ser acedida por aqueles que necessitam do conhecimento nela registado.

Um sistema de conhecimento tem que possuir, pois, ferramentas para aquisição do conhecimento, sua formulação e registo.

O termo **aquisição do conhecimento** refere-se a uma actividade onde são usadas técnicas que permitam que os computadores obtenham o conhecimento necessário ao sistema que se pretende construir.

A **formulação** do conhecimento adquirido é a sua descrição e explicação do que ele significa.

Nestes sistemas, há uma preocupação na utilização de representações e métodos de raciocínio que tornem o conhecimento facilmente acessível.

Classicamente, um sistema de conhecimento tem a arquitectura apresentada na Figura 1.2.

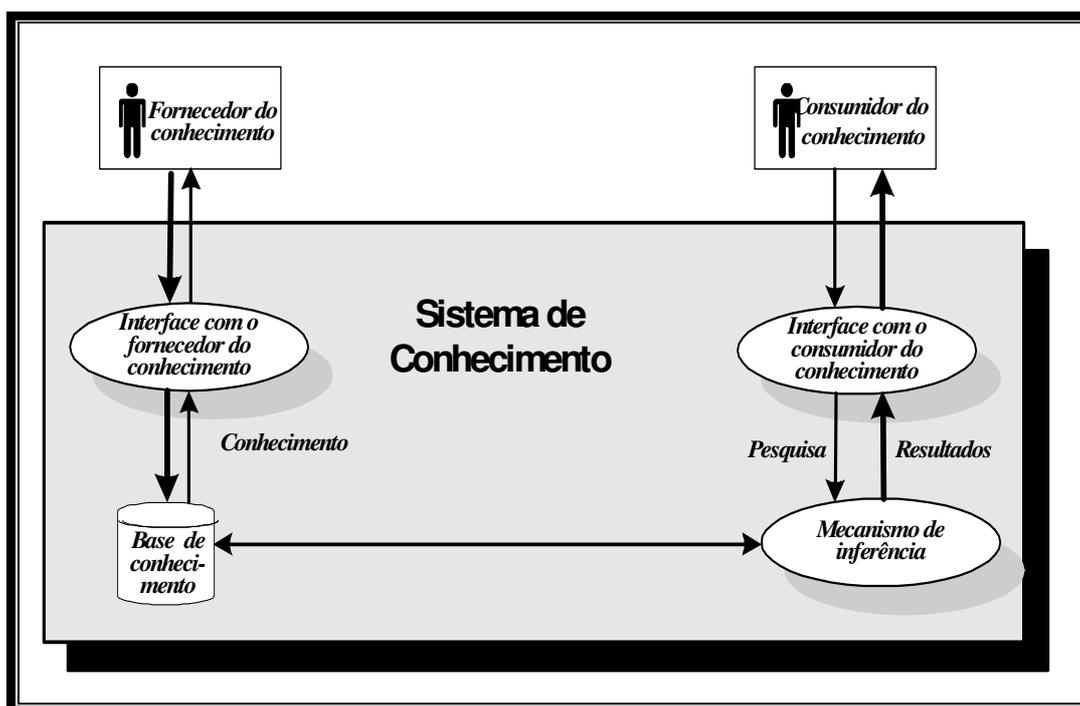


Figura 1.2 – Arquitectura dum sistema de conhecimento (Adaptado de [Stefix, 1995]).

A base de conhecimento é o repositório onde é registado o conhecimento adquirido pelo sistema e deve ser actualizada periodicamente a fim de que o seu conteúdo reflecta a evolução sofrida pelo conhecimento.

O *interface* com o fornecedor do conhecimento (e.g., DCBD, perito do negócio) permite o registo do conhecimento através de uma interacção máquina/máquina ou máquina/homem.

O *interface* com o consumidor do conhecimento (e.g., gestor, aplicação informática) é a parte do sistema que permite que o consumidor (homem ou sistema), através de um

mecanismo de inferência⁶, pesquise a base de conhecimento a fim de obter conhecimento aí registado ou outro a partir dele derivado.

Um sistema de conhecimento e a respectiva base de conhecimento, sendo sistemas computacionais, trazem vantagens sobre qualquer sistema manual de representação de conhecimento quer pela sua possibilidade de actualização em tempo real, quer pela sua facilidade de integração noutros sistemas ou adaptação imediata a novas necessidades.

Há questões que são pertinentes quando se pretende implementar um sistema de conhecimento numa organização. Uma reflexão atenta sobre essas questões e as respectivas respostas permite concluir da vantagem (ou não) da sua implantação:

- O que se pretende, na realidade, que faça o sistema de conhecimento?
- É necessário ou é, apenas, mais uma ferramenta?
- Quem tirará, na organização, proveito deste sistema?
- Quem o deve utilizar?
- Que classes de problemas pode resolver?
- Quais as suas limitações?
- Como se pode ligar a outras actividades para criar e usar conhecimento?

Assim, na construção dum sistema de conhecimento, algumas características e requisitos têm que ser tidos em conta, nomeadamente:

- o estabelecimento de limites no sistema de conhecimento a ser formalizado (e.g., num par (atributo,valor) limitar os atributos, seleccionando apenas alguns deles ou impor restrições aos possíveis valores, restringir o conhecimento a ser registado na base de conhecimento);

⁶ No âmbito deste trabalho, consideramos como *mecanismo de inferência* todo o mecanismo que permite, a partir de conhecimento existente, inferir novo conhecimento. Um mecanismo de inferência permite obter conclusões sobre propriedades não observadas de certas ocorrências, a partir de outras que foram observadas [Wand *et al.*, 1995].

- o tipo de representação escolhida (e.g., árvores de decisão, regras de produção, redes semânticas, lógica de predicados);
- a linguagem de desenvolvimento das ferramentas a utilizar (e.g., *Pascal*, *Prolog*).

Um protótipo do sistema ZACCAR, implementado em *Prolog*⁷ e usando regras de produção como formalismo de representação do conhecimento, foi testado em dois casos.

O protótipo foi implementado em ambiente *Windows NT*, usando um computador *Pentium III*; a versão do *Prolog* utilizada foi o *Win LPA-Prolog*⁸.

Este sistema pode ser aplicado quando a ferramenta de DM utilizada é uma ferramenta que gera padrões sob uma forma que seja compatível com regras.

A ferramenta de DM usada no estudo de casos foi o *Clementine* [Berson *et al.*, 1999], [*Clementine*, 1997]⁹.

O protótipo do sistema desenvolvido é suficientemente genérico, podendo ser utilizado nos mais variados domínios. Os estudos de caso levados a efeito situam-se no domínio comercial uma vez que neste domínio pode-se evidenciar, facilmente, o conceito desenvolvido e a evolução temporal dum padrão de comportamento é muito significativa quer para o desenvolvimento de campanhas de *marketing* quer para outras acções que têm como alvo os clientes.

1.2 Contribuições

As principais contribuições desta tese podem ser sintetizadas no seguinte conjunto:

⁷ O *Prolog (Programming in Logic)* caracteriza-se por ser uma linguagem baseada em cláusulas de Horn utilizando técnicas que combinam a resolução directa com o encaminhamento para trás (*backward chaining*) como procedimento de prova (dedução) [Clocksin, 1987], [Santos, 1999].

⁸ www.lpa.co.uk

⁹ www.spss.com/Clementine

- apresentação do conceito a “Gestão do CCC” que conduz a uma nova tarefa organizacional “Zelar pelo CCC”;
- estabelecimento de uma forma de estrutura do CCC e seu registo;
- concepção e exploração dum sistema de conhecimento para apoio à gestão do conhecimento do comportamento dos clientes - o sistema ZACCAR - cujo objectivo principal é permitir a viabilidade da nova tarefa através da:
 - ▶ recolha e uniformização dos padrões de comportamento obtidos com uma ferramenta de DM;
 - ▶ confrontação desses padrões com conhecimento já existente sobre o comportamento dos clientes, actualizando-o;
 - ▶ validação e documentação do conhecimento já actualizado feito pelo gestor do conhecimento organizacional;
 - ▶ integração do conhecimento depois de actualizado e completado, numa base de conhecimento que fará parte integrante do conhecimento organizacional.
- processo de consolidação do conhecimento descoberto em bases de dados, resolvendo problemas de interpretação e integração.

No sistema desenvolvido foi incorporado um módulo que permite a visualização do conhecimento pelo gestor do conhecimento organizacional, possibilitando-lhe a inferência de novo conhecimento.

Na prossecução dos objectivos presentes na elaboração deste trabalho, é feita uma análise pormenorizada da prática de CRM e sua relação com o conhecimento organizacional bem como do CCC com ênfase no tratamento que é dado a este conhecimento.

É evidente que as organizações, a fim de sobreviverem nos mercados actuais, adaptando-se, rapidamente, às novas exigências, necessitam de possuir meios para registar e utilizar o conhecimento sobre as suas actividades sejam elas resultantes dos contactos com os clientes ou outras.

As interações entre a empresa e os seus clientes são registadas, não interessando aos membros da organização onde nem como. No entanto, o conhecimento a partir delas obtido tem que se encontrar a um nível de fácil e imediato acesso aos consumidores deste conhecimento, sejam eles pessoas ou aplicações.

Com o sistema proposto, as organizações podem passar a dispor de uma base de conhecimento sobre o CCC, actualizada dum forma semi-automática onde está ainda registada uma evolução dos padrões de comportamento dos clientes e que faz parte integrante do conhecimento organizacional.

Assim, os beneficiários desta tese podem ser todas as organizações as quais, com fins embora diversos, têm que gerir o seu relacionamento com os clientes, precisando, para tal, de conhecer o CCC e a sua evolução.

É ainda preocupação deste trabalho a construção dum sistema fiável e aberto, no sentido em que permite a sua aplicação nos mais diversos domínios e a possibilidade de utilização, para obtenção dos padrões de comportamento dos clientes, de diferentes ferramentas ou algoritmos de DM.

A construção de um sistema que tire vantagens do conhecimento obtido com DM pode ser muito útil numa organização pois alguns padrões de comportamento que previamente eram muito difíceis ou mesmo impossíveis de registar, poder-se-ão tornar facilmente acessíveis, permitindo, em qualquer altura, a sua detecção.

As principais vantagens que o ZACCAR pode trazer para as organizações são:

- formalizar o CCC disperso na organização e guardá-lo, dum maneira semi-automática, num único repositório a partir de onde pode ser utilizado de maneira eficiente no negócio;
- permitir uma maior facilidade de comunicação entre os peritos de DCBD e os peritos de negócio facilitando, por esse motivo, quer o processo de DCBD quer a exploração do CCC;

- criar condições para que uma memória organizacional fique formalmente registada, independentemente dos gestores em cada momento da vida da empresa ou dos contextos em que as tarefas de DM foram executadas;
- permitir o desenvolvimento de novo conhecimento e a sua assimilação com conhecimento já existente.

Uma vantagem emergente da base de conhecimento criada pelo ZACCAR é que passa a ser possível melhorar os procedimentos habituais e, também, viabilizar novas práticas como sejam a possibilidade de aceder à eficácia de campanhas de *marketing* através do confronto do comportamento dos clientes destas acções com padrões de comportamento previamente detectados e que se encontram na BCCC (**B**ase de **C**onhecimento do **C**omportamento dos **C**lientes).

1.3 Processo de Investigação

Detectou-se uma situação que é apresentada como problemática, podendo ser melhorada: a “**Gestão do CCC**”. O contributo desta tese, para a resolução daquela situação, é a proposta de uma actividade – “**Zelar pelo CCC**”; para a sua viabilização, construímos uma ferramenta usando tecnologia existente.

O desenvolvimento da ferramenta foi feito usando uma abordagem de prototipagem.

O processo de investigação começou com uma revisão da literatura sobre CRM, tendo como principal objectivo detectar algumas características sobre o conhecimento dos clientes utilizado nesta prática do negócio, nomeadamente:

- como se obtém;
- como é registado;
- como é feita a consolidação do conhecimento já existente com o novo conhecimento.

Numa segunda fase da investigação, fizemos uma revisão da literatura na área das tecnologias para a obtenção e registo do conhecimento (bases de dados e *data warehouses*), mecanismos de inferência para transformar o conhecimento factual em conhecimento conceptual [Carvalho & Morais, 2001] (por exemplo, DM) e meios para o seu registo (bases de conhecimento).

Após estas fases de investigação que conduziram à detecção da situação problemática já referenciada, procedemos a uma análise cuidadosa do que realmente seria importante e não se encontrava desenvolvido: **Zelar pelo CCC**. Para viabilizar esta actividade, considerámos necessária a especificação e desenvolvimento dum sistema de conhecimento cujas características definimos, procedendo, posteriormente, à construção dum protótipo. Para a validação do protótipo construído, efectuámos dois estudos de casos [Cassell & Symon, 1997], [Cavaye, 1996].

Os dados para o primeiro estudo (Anexo 5) resultaram duma experimentação simulando uma situação real: em quatro momentos diferentes, foi solicitado, a alunos de três estabelecimentos de ensino superior, o preenchimento de um inquérito (Anexo 4) organizado por uma agência de viagens fictícia que propunha, em cada momento, uma viagem de fim de curso. Embora o destino fosse o mesmo, em cada momento a modalidade da viagem era diferente.

No segundo estudo, usou-se dados reais de uma empresa de indústria de confecção (Anexo 7). O estudo foi, como o anterior, dividido em momentos os quais correspondem a diferentes campanhas levadas a cabo pela empresa.

A análise dos resultados obtidos na utilização do protótipo permitiu-nos aferir da viabilidade do sistema proposto mas, também, a definição dum conjunto de novos requisitos tendo em vista o seu aperfeiçoamento em trabalho futuro.

1.4 Estrutura da Tese

A tese, além deste capítulo introdutório, apresenta mais 6 capítulos (Figura 1.3), um conjunto de anexos considerados de interesse quer para o trabalho desenvolvido quer para futuros melhoramentos que sobre ele possam vir a ser considerados, um índice de autores e um índice remissivo.

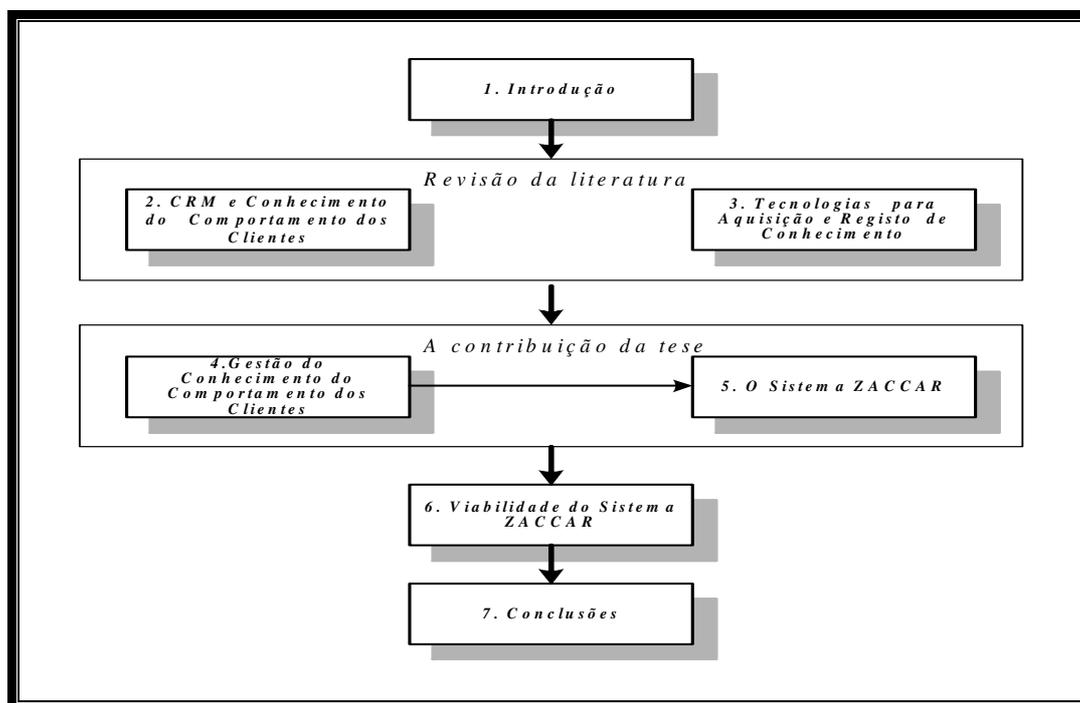


Figura 1.3 - Estrutura da tese.

Num primeiro grupo (Figura 1.3), intitulado "Revisão da literatura", composto pelos capítulos 2 e 3, temos a preocupação de contextualizar o trabalho desenvolvido quer analisando a posição das pessoas (ou áreas da organização) que poderão ser afectadas por ele quer fazendo uma exploração de tecnologia existente para a aquisição e registo do conhecimento. Consideramos que só assim é possível atribuir valor ao trabalho presente.

No capítulo 2 - **CRM e Conhecimento do Comportamento dos Clientes** - é feita uma abordagem à prática de CRM e são tecidas algumas considerações sobre o CCC, referindo a possibilidade da sua obtenção a partir do registo dos eventos cliente-empresa e o seu posicionamento no conhecimento organizacional.

No capítulo 3 – **Tecnologias para Aquisição e Registo de Conhecimento** - fazemos uma explanação sobre tecnologias existentes para a obtenção dos dados dos clientes (leitores de códigos de barras, caixas ATM (*Automatic Transaction Machines*), Internet, *e-mail*, *call centers*), suportes informáticos para o seu registo (bases de dados, *data warehouses* e *data marts*), a obtenção de padrões de comportamento dos clientes e modelos para a sua representação (redes neuronais, árvores de decisão e regras) e o registo destes padrões em bases de conhecimento.

A inclusão deste capítulo na tese tem como objectivo posicionar o leitor naquilo que existe e é, regularmente, utilizado nas empresas, para melhor evidenciar, neste processo, uma lacuna ainda existente: a Gestão do CCC.

Dum segundo grupo (Figura 1.3), "A contribuição da tese", fazem parte os capítulos 4 e 5 que pretendem apresentar ao leitor o trabalho desenvolvido e as principais contribuições desta tese.

O capítulo 4 é intitulado **Gestão do Conhecimento do Comportamento dos Clientes**. Este capítulo é dedicado à apresentação do grande objectivo da tese que é demonstrar o interesse, para as organizações, da “Gestão do CCC”. Assim, são apresentados o conceito “Gestão do CCC” e a nova actividade organizacional “Zelar pelo CCC” bem como o sistema de conhecimento necessário à sua viabilidade.

O capítulo 5 - **O Sistema ZACCAR** - é dedicado ao sistema proposto, descrevendo os requisitos técnicos inerentes à sua especificação, execução e implementação bem como algumas características do protótipo desenvolvido.

É dada particular importância ao processo de consolidação do conhecimento descoberto através da utilização de ferramentas de DM, uma funcionalidade chave do ZACCAR.

A exequibilidade do sistema é aferida no capítulo 6 - **Viabilidade do Sistema ZACCAR** - através da apresentação de duas experiências em que o protótipo do sistema foi testado. Com esta experimentação, procurámos encontrar os limites das capacidades do sistema.

Discute-se, ainda, os problemas com que nos deparámos no tratamento destes casos os quais abrem perspectivas futuras para o melhoramento do sistema desenvolvido.

No capítulo 7 - **Conclusões** - fazemos o fecho do trabalho realizado, apresentando as conclusões principais a que chegámos, discutindo as vantagens do sistema desenvolvido e os resultados obtidos bem como apresentando sugestões para desenvolvimentos futuros.

2. CRM e Conhecimento do Comportamento dos Clientes

*“The key to success in business is to know something that nobody else knows.”
(Aristotle Onassis¹⁰)*

A capacidade de atrair e conservar os clientes¹¹ é uma das chaves do sucesso no mundo dos negócios. Atrair clientes [Berson *et al.*, 1999] passa, por exemplo, por encontrar clientes que:

- previamente não são conhecedores dos produtos¹² que a empresa comercializa ou
- supostamente, não são candidatos à compra desses produtos mas poderão vir a sê-lo ou
- já compram esses produtos a empresas competidoras.

Para a retenção dos clientes, a empresa deve adaptar-se-lhes e antecipar as suas expectativas, oferecendo-lhes os produtos que eles desejam, num modo cativante e nas condições mais favoráveis através de um conjunto de práticas que, globalmente, tomam a designação de CRM (*Customer Relationship Management*) [Berson *et al.*, 1999], [Brown, 2000], [Newell, 2000a], [Burnett, 2001].

Os clientes não são iguais e, por isso, nem todos têm as mesmas preferências de aquisição de bens ou as mesmas exigências de serviços.

Mas, saber, em cada momento, o que os clientes realmente pretendem, passa pela identificação do cliente e pela aquisição de Conhecimento sobre o Comportamento dos Clientes (CCC), a sua conservação e, principalmente a sua gestão, i.e.:

¹⁰ Aristotle Socrates Onassis – Armador grego; viveu de 1906(?) a 1975

¹¹ *Cliente* – aquele que tem por objectivo adquirir um bem ou um serviço [Montcel, 1972].

- a sua confrontação com o conhecimento deste tipo já existente na organização;
- a actualização do conhecimento;
- o registo do conhecimento bem como da evolução temporal verificada.

Tal possibilidade requer a exploração das grandes quantidades de dados operacionais que as organizações registam diariamente e que estão a ser continuamente actualizadas.

2.1 A Prática de CRM

Nos últimos anos e como consequência de factores diversos que provocaram alterações nas regras dos negócios¹³, a função da empresa já não é impor os seus produtos mas sim vender os produtos que o cliente deseja.

É importante que as empresas possuam conhecimento específico sobre cada um dos seus clientes para que, em cada contacto cliente-empresa, tenham condições imediatas de escolher o produto adaptado à medida do cliente.

Por este motivo, o ponto de partida para que uma organização tire, efectivamente, vantagens de uma prática de CRM é a informação sobre quem são os seus clientes e quais as suas necessidades e aspirações.

Conhecer o cliente significa ser capaz de estabelecer um relacionamento com ele baseado nas suas necessidades e atendendo ao seu valor para a empresa; com esse fim, é necessário ter um sistema capaz de registar as novas informações sobre cada cliente sempre que elas se tornem disponíveis e, também, criar meios para difundir essas informações pela empresa.

¹² Um *produto* é entendido como um bem ou um serviço.

¹³ Além dum conjunto de factores políticos e económicos, também tiveram um grande peso nestas alterações as novas facilidades de comunicações, uma enorme explosão no domínio da informática (*hardware* e *software*) e a *Internet*.

Torna-se indispensável zelar pelo conhecimento sobre o comportamento dos clientes pelo que é necessário obter esse conhecimento, armazená-lo e actualizá-lo, fazendo um registo da sua evolução ao longo do tempo.

Uma empresa de pequena dimensão mantém um relacionamento individual com os seus clientes, sabendo o que cada um prefere, as formas de pagamento por que opta normalmente e outras características comportamentais importantes (e.g., quantidades adquiridas e épocas de maior procura de determinado produto).

No entanto, nas empresas que possuem centenas ou milhares de clientes, o contacto pessoal desvanece-se e, por consequência, já não é possível obter conhecimento ao nível individual. Deve-se, então, encontrar novos processos de relacionamento com os clientes para que, a partir do conhecimento dos seus padrões de comportamento, a empresa possa oferecer, em cada momento, o que eles realmente precisam ou o que é oportuno vender-lhes.

No passado, quantos mais clientes tinha a empresa mais difícil era saber o que cada um queria e, portanto, gerir o relacionamento com eles. Actualmente e, dado o avanço tecnológico entretanto conseguido, a situação é perfeitamente diferente; mais clientes representam mais interações cliente-empresa e, conseqüentemente, mais dados que conduzem à descoberta de padrões de comportamento com maior grau de confiança, potenciando o estabelecimento de uma relação cliente-empresa muito mais eficiente.

Pelos motivos expostos, um dos conceitos de gestão mais em destaque nas organizações actualmente é o de CRM.

Numa pesquisa em literatura da especialidade ([Berson *et al.*, 1999], [Brown, 2000], [Newell, 2000a], [Berry & Linoff, 2000], [Burnett, 2001], www.crm-forum.com, entre outros), encontram-se numerosas definições para CRM desde as que referem CRM como uma simples tecnologia às mais abrangentes que quase consideram que CRM é, por si só, o motor condutor da *BI*.

No contexto deste trabalho, definimos CRM como um processo que, baseado na alteração gradual do comportamento do cliente e na aprendizagem em cada interacção com ele, tem, como objectivo, um tratamento personalizado e o fortalecimento das relações cliente-empresa, tendo presente que há uma forte correlação entre a satisfação do cliente e a sua fidelização (Adaptado de [Newell, 2000]).

Deste modo, para implementar uma prática de CRM, a empresa deverá [Peppers & Rogers, 2000]:

- caracterizar o cliente, i.e., saber, na realidade, quem é o cliente e possuir informação detalhada sobre ele de tal modo que seja capaz de o identificar em qualquer ponto de contacto (e.g., pontos de venda, correio electrónico, *call centers*, páginas *Web* de acesso a comércio electrónico);
- diferenciar o cliente, i.e., adaptar a oferta a diferentes clientes de acordo com as suas preferências e o valor que eles representam para a empresa;
- interagir com o cliente procurando que as informações recolhidas contribuam para o aprofundamento da relação o que, muito provavelmente, diminui a probabilidade de afastamento do cliente;
- personalizar o cliente, i.e., a empresa terá de se adaptar para responder melhor às necessidades específicas do cliente por ele manifestadas durante a interacção o que consiste em tratar diferentes clientes de maneira diferente.

Em resumo, a empresa tem que chegar a um conhecimento tal dos seus clientes reais e potenciais que lhe permita agrupá-los, atendendo aos proveitos que eles possam trazer (Figura 2.1).

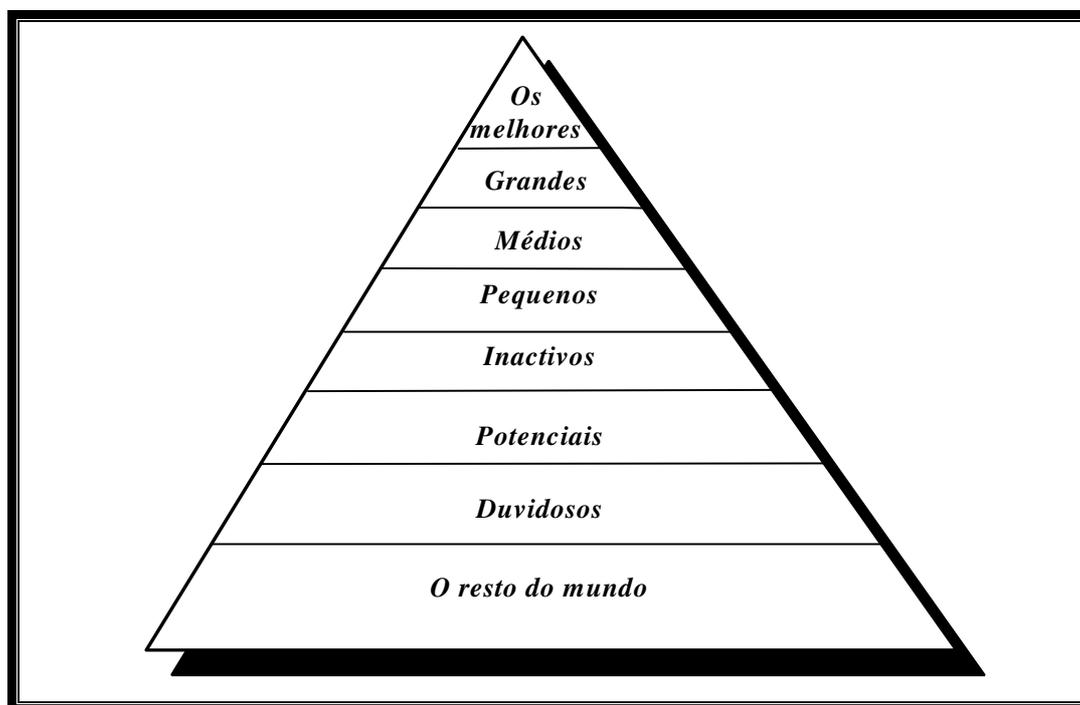


Figura 2.1 – Pirâmide de clientes reais e potenciais (Adaptado de [Curry & Curry, 2000]).

Apresenta-se, de seguida, alguns exemplos de aplicação, com sucesso, de uma prática de CRM:

Exemplo1: A cadeia de supermercados Dick's¹⁴ que tem oito lojas no estado de Wisconsin (U.S.A.), instalou um sistema - o *DataVantage*¹⁵ - cujo objectivo é prever quando os clientes voltarão a comprar determinados produtos; na posse desta informação, o sistema gera ofertas especiais.

O esquema de funcionamento é o seguinte: um cliente que gaste mais do que um montante previamente definido por semana, nesta cadeia de supermercados, recebe, ao fim de quinze dias, uma lista de compras personalizada, indicando, ainda, as ofertas a que tem direito nas próximas aquisições; a lista baseia-se no histórico de compras do cliente.

Quando o cliente chega à caixa registadora do supermercado, o operador faz uma leitura óptica do código de barras contido na lista de compras do

¹⁴ www.dickssupermarket.com

cliente. As ofertas que constam da lista são, automaticamente, descontadas na conta a pagar pelo cliente e o histórico das compras é actualizado a fim de ser elaborada a próxima lista.

A empresa personaliza as promoções, criando uma oferta directamente proporcional ao valor do negócio do cliente e aos seus interesses. Assim, por exemplo, um cliente que tem um animal de estimação tem direito a descontos na comida do animal ou, se tiver um bebé, é contemplado com uma promoção nas fraldas ou outros produtos infantis.

Com este tratamento pela parte da empresa, o cliente apercebe-se de que se fizer compras noutra supermercado perde algumas das ofertas que são geradas pelo histórico das suas compras; daí tem todo o interesse em se manter fiel àquela cadeia de supermercados [Peppers & Rogers, 2000].

Exemplo2: Os armazéns *Hallmark Gold Grown*¹⁶ representam uma das maiores cadeias de venda a retalho nos Estados Unidos, com 5000 lojas abertas ao público; entre os artigos que comercializam, destaca-se os cartões de felicitações.

Esta empresa iniciou, em 1994, um programa de CRM, conhecido como *Global Grown Card* que teve por base o lançamento dum cartão de fidelidade a poder ser usado pelos clientes em qualquer uma das lojas da empresa.

Com o uso do cartão, os clientes ganham pontos. Por sua vez, a empresa usa os dados registados com o programa para aprender o que cada cliente vale nesta relação, que produtos têm valor para ele e o que diferencia cada cliente dos outros.

Além dos pontos obtidos que dão ao cliente direito a brindes ou promoções, a *Hallmark* segmenta os seus clientes de acordo com o

¹⁵ www.datavantage.com

¹⁶ www.hallmark.com

conhecimento obtido nesta relação e, aos melhores de entre eles, dá um tratamento especial.

É mantido um contacto directo com os clientes quer por *e-mail* quer por outras vias, nomeadamente a realização de reuniões por grupos em que são apresentados os novos produtos e promoções e solicitadas sugestões. É, ainda, mantido um serviço especial que tem muito a ver com a comercialização dos cartões e que consiste na manutenção dum ficheiro com os nomes, direcções e datas importantes relativamente aos familiares e amigos do cliente, alertando este sobre a altura de envio dum cartão de felicitações.

Sendo a *Hallmark* formada por uma cadeia de lojas espalhadas pelos Estados Unidos, além destes programas gerais, há a possibilidade de, em cada loja, dado esta ter acesso em tempo real ao conhecimento sobre os seus clientes directos, serem realizados eventos locais para atrair cada vez mais clientes e fidelizar os já existentes [Newell, 2000a].

Uma empresa que pretende implantar CRM com sucesso deve seguir algumas práticas fundamentais, nomeadamente:

1. segmentar os seus clientes¹⁷ de acordo com perfis de comportamento;
2. investir em tecnologia adequada;

Podemos, então, facilmente concluir que o sucesso de uma prática de CRM depende, em grande parte, da análise que é realizada sobre o conhecimento do comportamento dos clientes, análise essa que deve ter em conta o carácter dinâmico do conhecimento; tal processo leva a que a empresa fique a saber, por exemplo, de entre os produtos que pretende vender, aqueles que o cliente realmente quer comprar, os meios usados para a

¹⁷ *Segmentar clientes* consiste em agrupar clientes de acordo com o relacionamento que eles mantêm com a empresa. Esta prática permite, entre outros, a possibilidade de identificar os grupos de clientes que representam o maior proveito para a empresa [Curry & Curry, 2000].

sua aquisição, as épocas preferenciais para as compras, as associações de produtos¹⁸ e os clientes que são ou não sensíveis a determinado tipo de campanhas.

A análise do comportamento dos clientes tem sempre como objectivo, pelo menos, um dos pontos seguintes:

- retenção dos clientes leais à empresa;
- angariação de novos clientes;
- diminuição dos atritos com os clientes;
- venda de produtos diferentes ao mesmo cliente;
- venda de maior quantidade dum produto a determinado cliente.

O maior proveito de CRM – uma das maiores preocupações da empresa - é, sem dúvida, a retenção dos clientes leais à empresa pois são estes que têm tendência para adquirir mais produtos ou requisitar mais serviços com a vantagem de que não trazem, com eles, custos adicionais de angariação; para esta retenção contribui, concerteza, o grau de satisfação do cliente o qual aumenta com o tratamento individualizado que a empresa lhe der.

Toda a interacção com um cliente é uma fonte potencial de informação que permite aumentar o conhecimento sobre o cliente, contribuindo, provavelmente, para o estabelecimento de uma relação duradoura da qual ambas as partes – cliente e empresa – só poderão vir a beneficiar. Com este conhecimento, a empresa obtém vantagens no sentido em que pode passar a oferecer, ao cliente, produtos à medida das suas necessidades e desejos.

CRM pode conduzir a campanhas de *marketing* baseadas em segmentos específicos de clientes ou mesmo em clientes individuais; uma das áreas a explorar para a efectuação de uma dessas campanhas é a análise de eventos relacionados com clientes.

¹⁸ Os **produtos** dizem-se **associados** se têm ocorrências em simultâneo nos registos das bases de dados.

Ao longo da vida de cada indivíduo, ocorrem eventos que, se conhecidos, podem conduzir a iniciativas proveitosas da parte das organizações. Alguns exemplos são apresentados na Tabela 2.1.

Tabela 2.1 - Eventos na vida dum cliente e iniciativas de CRM que podem desencadear.

<i>Evento</i>	<i>Iniciativas possíveis</i>
Mudança de casa	<ul style="list-style-type: none"> - uma companhia telefónica pode contactar o cliente para lhe ceder um telemóvel durante o período de mudança; - um banco pode propor uma carteira de empréstimos para que o cliente possa fazer face a novas necessidades (e.g., aquisição de electrodomésticos, mobílias); - uma companhia de seguros pode enviar uma proposta de seguros tendo em atenção as novas instalações; - uma empresa de venda a retalho pode enviar informação sobre a localização do seu centro de venda mais próximo da nova morada do potencial cliente.
Nascimento dum filho	<ul style="list-style-type: none"> - as empresas de venda a retalho podem enviar catálogos de produtos úteis nessa ocasião quer para a casa quer para o bebé e, até mesmo, informações sobre promoções apropriadas ao momento; - um banco pode sugerir um plano financeiro para a criança; - as companhias de aviação podem oferecer <i>check-ins</i> especiais para o futuro utilizador; - uma companhia de seguros pode apresentar um plano de seguros para o potencial novo cliente.
Férias	<ul style="list-style-type: none"> - uma empresa telefónica pode propor a aquisição de cartões de crédito telefónico; - companhias de seguros podem oferecer seguros de viagem; - um banco pode propor condições especiais para as deslocações ao estrangeiro; - empresas de segurança podem propor um conjunto de alarmes e outras medidas apropriados à situação.

Outros eventos são, também, dignos de atenção podendo ser motivo para que as empresas efectuem campanhas especiais, nomeadamente:

- Uma chamada telefónica dum desconhecido para o *call center* de uma empresa a pedir informações sobre um produto que esta comercializa pode levar a que a empresa o procure directamente ou através do envio de catálogos, prestando informações mais detalhadas sobre o produto em causa e apresentando outros produtos afins.
- Um cliente dum supermercado que, de repente, deixa de fazer ali as suas compras; o estabelecimento tentará detectar as razões que conduziram a tal mudança e

procurará processos (por exemplo, promoções ou oferta de novos serviços) para recuperar o cliente se o considerar como um cliente proveitoso.

- Uma conta bancária cujo saldo sofre um grande acréscimo relativamente ao que é habitual desencadeará um conjunto de acções do banco, que se manifestará através da apresentação de novas modalidades de depósitos favoráveis para o cliente.
- Um cliente dum cartão de crédito que altera o seu modo de pagamento mensal; a instituição de crédito após tentar analisar as razões que conduziram a tal atitude, contactará o cliente, sugerindo-lhe outros processos de pagamento que, eventualmente, poderão ser mais favoráveis a este.
- Um cartão de crédito que passa a ser utilizado em aquisições que não fazem parte do histórico do seu possuidor. Este é um evento que deve conduzir, de imediato, a instituição de crédito a uma posição de alerta pois tal atitude pode traduzir uso fraudulento do cartão.

Não é difícil, com o actual desenvolvimento tecnológico, obter informação sobre estes eventos ou outros; basta, no geral, explorar os registos das transacções (compras, consultas, reclamações) que os clientes tiveram com a empresa. Módulos de *software*, conhecidos como *triggers*¹⁹, são executados automaticamente pelos sistemas informáticos quando estes detectam alterações pré-definidas nas bases de dados, podendo conduzir, de imediato, a acções no âmbito de CRM [Silberschatz *et al.*, 1996], [Berson *et al.*, 1999].

Tomadas de atitude das empresas face aos eventos referidos levam a que os clientes se apercebam de que elas oferecem todas as condições para lhes dar um tratamento individualizado; assim, poderão reconhecer os benefícios, para ambas as partes, do estabelecimento de uma relação estável.

¹⁹ *Trigger* é um programa informático que é activado quando uma dada situação é detectada.

Por último, para que uma prática de CRM tenha sucesso, um conjunto de pressupostos devem ser tidos em conta numa organização:

- a informação sobre uma transacção ou uma solicitação de serviço efectuada através dum qualquer ponto de contacto deve, de imediato, ser integrada num sistema acessível a todos os canais da empresa;
- toda a informação histórica sobre os clientes, incluindo a sua evolução temporal, deve ser disponibilizada, atempadamente, aos responsáveis dos serviços;
- devem ser implementados processos de quantificação:
 - ▶ do proveito obtido com cada cliente;
 - ▶ da eficácia de campanhas de *marketing*;
- o conhecimento sobre os clientes deve ser tido em conta em estratégias de venda e campanhas de *marketing*;
- devem existir programas específicos para premiar a fidelidade dos clientes;
- os clientes devem ser contactados pela organização através do seu canal preferencial (e.g., telefone, correio electrónico, *Internet*, *fax*, pessoalmente);
- deve existir uma cultura de empresa baseada não no produto mas no cliente;
- os clientes reais ou potenciais devem estar segmentados de acordo com as suas preferências de aquisição de bens ou necessidade de serviços;
- a informação sobre os clientes deve ser gerida independentemente do canal utilizado e da sua localização.

2.2 O Cliente

Como se viu anteriormente, para que se possa estabelecer uma conveniente prática de CRM é necessário:

- identificar o cliente do negócio;
- identificar as várias acções executadas pelo cliente;
- analisar a evolução, ao longo do tempo, da relação cliente-empresa.

A identificação do cliente, embora, à primeira vista, possa parecer demasiado evidente não o é tanto assim se se analisar a questão com mais pormenor [Berry & Linoff, 2000].

Vejamos, alguns exemplos que melhor podem exprimir esta ideia:

Exemplo 1: Para uma instituição bancária que pretende fazer uma análise do histórico de diferentes tipos de contas, o cliente corresponde a uma conta individualmente, independentemente do facto de haver indivíduos que possuam várias contas.

No entanto, se a intenção for fazer uma análise a fim de, por exemplo, encontrar segmentos de clientes tendo em vista uma campanha para o lançamento de novos produtos financeiros, então o cliente já é o indivíduo associado à totalidade das suas contas.

Exemplo 2: Numa cadeia de supermercados, se o interesse do negócio for o estudo de comportamento com cartões de fidelidade, o cliente é o indivíduo identificado pelo seu número de cartão.

Mas, se o interesse for uma análise de associação de compras a fim de, por exemplo, estudar a melhor distribuição dos produtos nas prateleiras do supermercado, então o cliente é cada um dos conjuntos de transacções efectuadas por um mesmo indivíduo, numa determinada data e local.

Exemplo 3: Para uma empresa bancária que forneça cartões de crédito, um cliente pode ser:

- o indivíduo possuidor de uma conta, independentemente dos cartões de crédito que lhe foram atribuídos;
- uma empresa que é responsável por todos os cartões de crédito que distribui a alguns dos seus funcionários;
- um cartão de crédito.

Há, assim, que analisar, atentamente, o negócio em si para, então, depois, poder definir quem, na realidade, é o cliente desse negócio o que é essencial para que se

possam construir padrões de comportamento e efectuar acções tendo em vista a oferta de melhores serviços ou a execução de campanhas de *marketing* mais eficientes as quais pretendem conduzir quer a um maior lucro do negócio quer a um relacionamento mais eficiente com o cliente [Newell, 2000b].

Todos os indivíduos, como seres humanos que são, possuem um ciclo de vida ao longo do qual podem ocorrer eventos variados como sejam a obtenção dum emprego, o casamento ou nascimento dum filho. Estes eventos geram necessidades de aquisições de produtos ou de requisição de serviços diferentes, representando, por esse motivo, uma importante fonte de informação para as organizações.

Entre um indivíduo e a empresa da qual é cliente, estabelece-se, também, um ciclo de vida (Figura 2.2) composto pelas possíveis fases de relacionamento entre ambos que evolui ao longo do tempo e o conhecimento desta evolução é fundamental para a compreensão do comportamento do cliente e, conseqüentemente, para acções no âmbito de CRM.

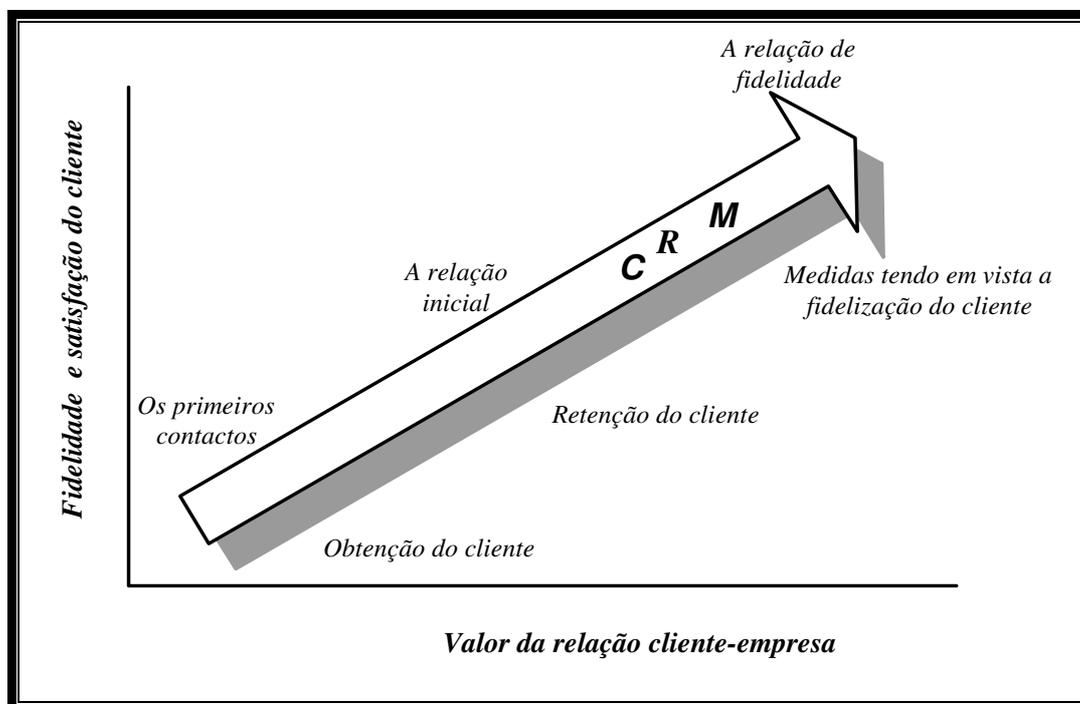


Figura 2.2 – Ciclo de vida cliente-empresa (Adaptado de [Brown, 2000]).

Exemplifiquemos este ciclo através da utilização dum portal de uma empresa que pratica comércio electrónico:

1. Numa primeira fase, o cliente é apenas um potencial cliente, i.e., alguém que visitou o portal da empresa, se calhar mais do que uma vez, pondo, eventualmente, questões que mostram o seu interesse pelos produtos que estão a ser comercializados.

O conhecimento sobre este cliente é extremamente vago, limitando-se a registos gerais no ficheiro de *logs*²⁰ do sistema informático da empresa.

2. Na fase seguinte, o potencial cliente inscreve-se como cliente da empresa, preenchendo os formulários que esta disponibiliza na sua página. Os dados registados nesses formulários têm, como objectivo, a inclusão do cliente em grupos de acordo com necessidades ou interesses semelhantes a outros clientes já existentes.

3. Numa terceira fase, o cliente é já um cliente efectivo; está registado e, regularmente, faz aquisições de produtos da empresa em causa, ficando estes movimentos registados nas bases de dados da empresa.

A análise do comportamento de compras deste cliente e a sua evolução podem levar ao estabelecimento de conclusões sobre o seu comportamento futuro ou sobre o de potenciais clientes com perfis semelhantes a ele.

4. Numa quarta fase, que pode ou não ocorrer, o cliente, voluntária ou involuntariamente, deixa de aceder ao portal da empresa e, conseqüentemente, não volta a comprar os produtos que ela comercializa.

A análise dos motivos para este abandono, poderá ser muito útil à empresa a fim de estar prevenida contra futuros abandonos de clientes com perfis idênticos a este.

²⁰ **Log** - ficheiro onde ficam registados todos os acessos ao sistema, incluindo os dos clientes ou potenciais clientes da empresa.

Atendendo a que a preocupação da empresa é, fundamentalmente, a conservação dos seus clientes mais leais mas também a aquisição de novos clientes com perfis iguais àqueles, campanhas de *marketing* com características diferentes devem ser levadas a cabo nas diferentes fases do ciclo de vida do cliente.

2.3 O Conhecimento sobre o Comportamento dos Clientes (CCC)

Independentemente do estatuto e da profissão, todos são potenciais clientes e todos fornecem conhecimento às empresas nas mais diversas acções ao longo da vida [Blackwell *et al.*, 2001]. Sempre que efectuamos uma chamada telefónica, usamos um cartão de crédito, apresentamos um cartão de cliente ou consultamos uma página na *Internet*, estamos a fornecer, às empresas, elementos que potenciam futuros estudos sobre o comportamento dos seus clientes.

Estes elementos vão permitir que as empresas possam analisar tendências, lacunas ou mudanças no comportamento de compras provocadas por factores diversos como sejam campanhas de *marketing*, podendo, então, ajustar a sua acção por forma a explorar esse conhecimento em benefício próprio.

Uma melhor compreensão, por exemplo, das necessidades dos clientes pode ajudar as empresas a não só tomar medidas preventivas contra o afastamento dos clientes já existentes mas também a atrair novos clientes.

Atendendo ao que já foi exposto neste capítulo, faz sentido considerar como CCC:

- os hábitos de consumo como sejam:
 - ▶ que produtos compram habitualmente;
 - ▶ em que quantidade adquirem cada produto;
 - ▶ periodicidade da compra de determinado produto;
 - ▶ combinações de produtos comprados, normalmente, em conjunto;
- os locais preferenciais para aquisição de determinados produtos;

- ocasiões escolhidas para efectuar determinado tipo de compras:
 - ▶ horas preferenciais ao longo do dia;
 - ▶ dias preferidos durante a semana;
 - ▶ períodos mais frequentes do ano;
- a distribuição geográfica de grupos de clientes;
- os canais preferenciais para os clientes contactarem a organização;
- o impacto que as campanhas de *marketing* ou outras iniciativas de contactos têm sobre os clientes;
- a evolução da satisfação dos clientes.

O CCC caracteriza detalhadamente o cliente, permitindo responder a questões como “Quem?”, “Onde?”, “O quê?”, “Quando?”, “Quanto?”, “Porquê?” (Figura 2.3).

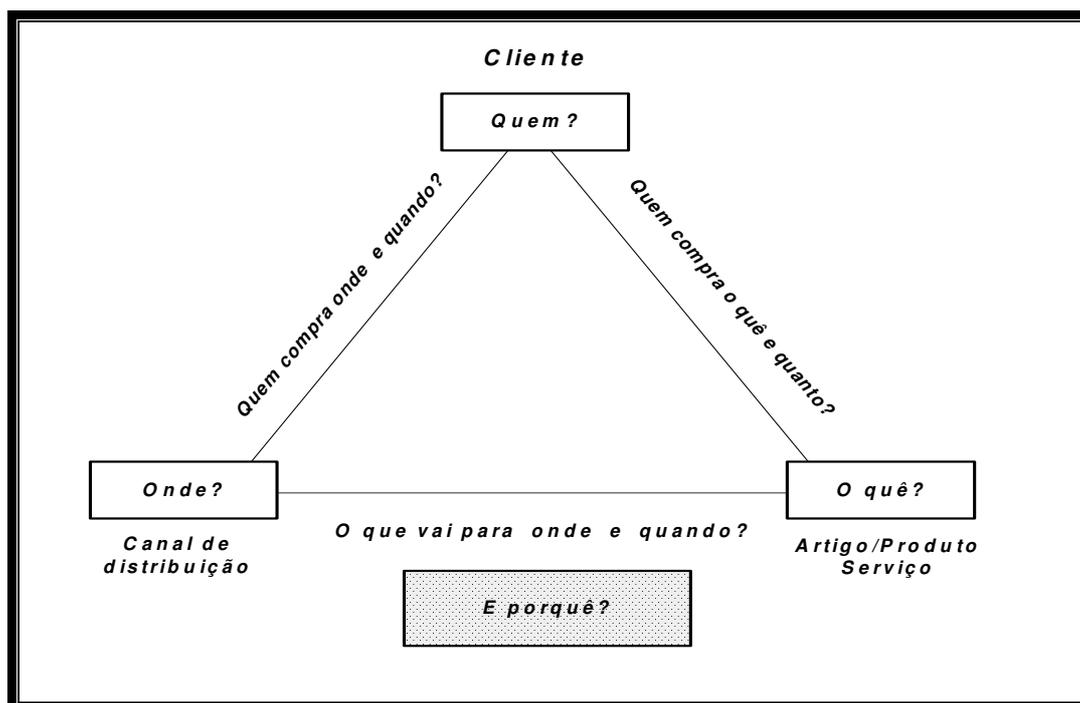


Figura 2.3 - Questões fundamentais para o conhecimento do comportamento dos clientes [Berson *et al.*, 1999].

Como tem vindo a ser referido, o comportamento do cliente evolui ao longo do seu ciclo de vida e factores externos à empresa ou certos eventos podem conduzir a alterações nos padrões de comportamento dos clientes. As justificações para essas alte-

rações devem ser conhecidas e ficar registradas, permitindo a realização de análises posteriores.

2.4 O CCC nas Empresas

O impacto potencial do CCC numa empresa varia com a dimensão desta, com a maneira como está organizada e com o tipo de produtos que disponibiliza aos seus clientes.

Numa grande empresa, o CCC pode ser mais problemático do que numa pequena empresa, na medida em que, nesta, há uma relação mais personalizada cliente-empresa e esta conhece bem os seus clientes, os seus hábitos de consumo e os processos para os cativar.

Por outro lado, nas grandes empresas, nomeadamente em áreas como a distribuição, a banca ou os seguros, o CCC assume grande relevância pois é dele que vai depender o crescimento da empresa e a sua quota de mercado, permitindo-lhe, em tempo útil, dar a melhor satisfação aos seus clientes.

O CCC é necessário para:

- predizer qual o perfil dos clientes que responderão a um *mail* específico, reagirão a uma campanha de *marketing* ou comprarão determinados produtos;
- verificar a eficácia de acções dirigidas à mudança do comportamento dos clientes.

Acções diversas na organização como sejam as campanhas de *marketing*, serão muito mais eficientes se forem planeadas tendo em consideração um conhecimento aprofundado sobre os clientes a quem essas campanhas se dirigem. Assim, um profissional de *marketing* deve: [Brito, 2000]

- identificar e diferenciar os clientes;
- implementar políticas de *marketing* individualizado;
- fidelizar clientes.

Uma boa campanha de *marketing* consiste em seleccionar uma classe de clientes e, tendo em conta certas características e restrições impostas à campanha, determinar que ofertas devem ser enviadas a que clientes, através de que canais e qual é a melhor altura para o fazer.

Estas campanhas podem ter em vista a obtenção imediata dum maior lucro para a empresa, resultante dum aumento de vendas mas, também, o estabelecimento dum melhor relacionamento cliente-empresa.

O CCC pode trazer os mais diversos benefícios para a empresa como sejam:

- A obtenção de elementos para comercialização de novos produtos ou serviços.

Exemplo: A companhia americana de seguros *GEICO*²¹ (*Government Employees Insurance Company*), tendo identificado os clientes de mais baixo risco, baseada nos dados que regista sobre os seus segurados, oferece-lhes seguros 15 a 20% mais baratos do que aqueles que as empresas competidoras comercializam.

- A determinação de mudanças nos hábitos de compra ou de utilização de serviços o que pode conduzir à necessidade de introduzir inovações na empresa.

Exemplo: A *Boots*²², uma empresa do Reino Unido, fundada em 1887, possui, actualmente, cerca de 1300 lojas que se dedicam ao comércio retalhista de produtos de saúde e beleza. O grupo de gestão da *Boots* reconheceu, recentemente, que, face ao mercado existente, havia que tomar uma iniciativa de aproximação aos clientes tendo em vista:

- ▶ uma melhor perspectiva dos clientes;
- ▶ a construção de uma base de dados que lhe dê a possibilidade de criar a oferta à medida das necessidades dos clientes;

²¹ www.geico.com

²² www.boots-plc.com

▶ a definição de um sistema de vendas baseado na fidelidade dos seus clientes.

A iniciativa tomada foi a criação dum cartão de fidelidade – o *Advantage Card* – que os clientes passaram a utilizar em qualquer das transacções efectuadas nas lojas desta companhia.

Com o conhecimento sobre os clientes adquirido como resultado da utilização do *Advantage Card*, a *Boots* obteve elementos para a oferta de novos produtos e serviços que aquele conhecimento indicou que os consumidores apreciariam.

Deste modo, esta empresa tornou-se líder em seguros de saúde e de viagem e entrou no mercado de venda directa com programas como o “*Mother and Baby at Home*” [Newell, 2000a].

- A definição de segmentos de mercado onde os produtos da empresa são usados, aqueles onde não o são e aqueles que apresentam todas as condições para que o sejam.

Exemplo: Na Caixa Geral dos Depósitos, as direcções de *marketing*, começaram a sentir a necessidade de conhecer melhor os seus clientes e de fazer campanhas mais eficazes.

Iniciaram, então, um projecto com o objectivo de aumentar a carteira de clientes com cartão de crédito o que passa por tentar descobrir quais os clientes com mais propensão para adquirir este tipo de cartão, num universo de cerca de 7,5 milhões de clientes.

Este projecto foi iniciado em 1999, na sequência da criação de uma plataforma de suporte de informação para gestão²³.

Durante o decurso do projecto, foi possível identificar padrões de comportamento de clientes que conduzem à detecção de quais os clientes que poderão estar mais interessados na aquisição de um cartão de crédito.

²³ www.fbnet.pt/red/0300/a03-00-00.shtml

- Orientações para a redefinição de páginas *Web* ligadas ao comércio electrónico ou à publicidade de produtos e serviços de forma a cativar mais clientes quer pelo seu aspecto apelativo quer pela sua facilidade de manutenção²⁴.

Exemplo: A *Amazon.com*²⁵, uma das maiores livrarias electrónicas existentes, investiu, fortemente, numa base de dados de informação livreira e em boas ferramentas de pesquisa.

Dada a qualidade da informação que através deste portal se pode obter, é frequente que muitos sejam os que a consultam antes de adquirir um livro de determinada área.

Uma das hipóteses de pesquisa é a indicação do nome dum determinado livro; quando tal acontece, aparece, também, uma relação de outros livros ou revistas dentro do mesmo assunto.

A par das ferramentas citadas, a *Amazon.com* possui, ainda, uma base de dados onde regista as consultas e aquisições efectuadas pelos seus clientes; este repositório permite-lhe, por exemplo, perante dois clientes que solicitam informação sobre o mesmo tema, apresentar páginas com um tratamento diferente, encaminhando-os de acordo com as suas preferências.

Um dos processos utilizados para a recomendação de livros é baseado na informação sobre os livros mais frequentemente comprados por clientes que também compraram o livro que está a ser pesquisado no momento.

Com este sistema, o próprio cliente tem interesse em ser fiel à *Amazon* pois, se for adquirir livros noutra lado, a livraria perde este conhecimento e os seus próximos conselhos serão menos eficazes.

- O estabelecimento dum relacionamento cliente-empresa personalizado.

²⁴ "If I have 3 million customers on the Web, I should have 3 million stores on the Web", Jeff Bezos, CEO of Amazon.com, apud [Kohavi & Provost, 2000].

²⁵ www.amazon.com

Exemplo: O *First Direct*²⁶, um banco directo do Reino Unido, possui um sistema informático com acesso a uma base de dados sobre os seus clientes o que permite, aos seus funcionários, durante uma chamada telefónica dum cliente, aceder a toda a informação relevante existente sobre esse cliente, iniciando-se, assim, um diálogo personalizado e eficiente o que, provavelmente, o cativará.

Pode-se considerar que, com as medidas que toma, este banco segue já uma intensa prática de CRM.

- O conhecimento dos clientes leais à empresa.

Exemplo: A *American Airlines*²⁷ desenvolveu um programa, chamado *Aadvantage* que regista os comportamentos dos seus clientes habituais. Através deste programa, cada utilizador pode-se tornar um membro *Aadvantage*, acumulando distâncias percorridas em milhas, tendo em vista obter bónus de viagens e outros benefícios.

No acto de inscrição ou contactos posteriores (e.g., marcações de viagens, consultas), o cliente fornece, directamente ou não, um conjunto de elementos sobre si e as suas preferências de viagens (por exemplo, regiões preferidas, duração de viagens, escolha de lugar no avião, número de viagens por ano, épocas de viagens) que ficam registadas na base de dados da companhia e que esta depois explora para seu proveito.

Com os elementos que regista, segmenta os seus clientes por padrões de viagens frequentes, oferecendo-lhes de acordo com o nível em que os classificaram, bónus em viagens ou enviando-lhes certificados que estes depois apresentam aos balcões da companhia, recebendo ofertas [Brown, 2000].

²⁶ www.firstdirect.co.uk

²⁷ www.aa.com

- A detecção de clientes envolvidos em transacções que incluem vários tipos de riscos tais como liquidez e fraude.

Exemplo: A empresa *Bits & Pixels*²⁸ desenvolveu uma tecnologia que permite "ver" clientes electrónicos. Com esta tecnologia, através da utilização de métodos de obtenção de padrões de comportamento, é possível detectar fraudes electrónicas, nomeadamente, fraudes bancárias.

Assim, a existência do CCC poderá contribuir para uma vantagem competitiva da organização e representa um importante recurso estratégico.

2.5 Razões para o Aumento da Importância do CCC

Periodicamente, a sociedade altera a sua visão do mundo, os seus valores básicos, a sua estrutura social e política e as suas instituições fundamentais. Todas estas mudanças têm impacto em áreas diversas, nomeadamente no negócio e nas regras que o conduzem.

Um exemplo destas mudanças é uma alteração recente do equilíbrio da balança de poder no negócio do vendedor que tinha toda a informação sobre os produtos e alternativas competitivas para o consumidor que tem, agora, a capacidade de analisar, comparar e escolher.

Há alguns anos atrás, a relação cliente-empresa era bem diferente da actual dadas as limitações existentes.

Os clientes não tinham, dada a pouca informação a que acediam, grandes alternativas de escolha, fazendo as compras para satisfazer as suas necessidades e usando a razão para determinar que produto ou serviço melhor se ajustava aos seus interesses o que não

²⁸ www.bitpix.com/cia/fraud.htm

é a prática actual em que factores de natureza muito diversa influenciam os hábitos de compra dos clientes.

Daquele modo, as empresas mantinham os mesmos clientes ao longo da sua existência, não lhes sendo preciso, para isso, grandes esforços. A organização tinha um produto que satisfazia as necessidades do cliente e este adquiria-o quando necessário. Poderia haver um bom relacionamento cliente-empresa mas não era exigido a esta um grande empenhamento para reter os seus clientes.

Os padrões de compra e de comportamento mantinham-se mais ou menos inalterados durante longos intervalos de tempo porque os principais determinantes do comportamento – preferências, percepções e disponibilidade de escolhas – não variavam com muita frequência [Cabena *et al.*, 1997].

Actualmente, o problema já não se apresenta da mesma maneira. A oferta do mercado é diversa e apelativa, os meios para chegar aos clientes são os mais variados que se possam imaginar, o próprio cliente é cada vez mais exigente e a situação da organização torna-se complicada se não mudar a sua maneira de ser, arranjando processos para cativar cada vez mais clientes ao mesmo tempo que fideliza os já existentes.

As organizações actuais vivem e viverão, cada vez mais, num ambiente onde os mercados estão, continuamente, a mudar, a tecnologia prolifera, os competidores multiplicam-se e os produtos tornam-se obsoletos rapidamente [Nonaka & Takeuchi, 1995].

Há fenómenos que surgiram, actualmente, como sejam a globalização gerada pelo comércio electrónico, a rápida evolução tecnológica, a individualização da sociedade e novos valores que são responsáveis por esta mudança [Krogh *et al.*, 1999].

Estas mudanças fundamentais no ambiente do negócio conduzem a alterações na maneira como as organizações vêem os seus clientes e planeiam aproximar-se deles. Entre estas mudanças, destacamos [Cabena *et al.*, 1997]:

- os consumidores têm acesso a uma melhor informação sobre os produtos, através de guias de compras e serviços, disponibilizados em catálogos que estão a ser frequentemente enviados para os seus domicílios, oferecidos gratuitamente em diversas ocasiões e locais ou disponíveis *on-line*; esta situação conduz a uma maior comodidade da parte do cliente quer na escolha dos bens ou serviços a

adquirir quer na sua solicitação a qual se pode processar por correio normal ou por encomenda electrónica;

- diariamente, somos abordados com ofertas de novos produtos que nos são facilitados em condições cada vez mais vantajosas;
- a competição aumenta e, conseqüentemente, os riscos do negócio;
- a *Internet* e o comércio electrónico permitem que uma empresa “invada” o mercado dos seus concorrentes mesmo que esteja a actuar no outro lado do mundo;
- com o advento da *Internet* e do comércio electrónico, a vantagem da localização do negócio quase desapareceu e este passa a depender muito mais do valor da informação que disponibiliza aos seus clientes [Schafer *et al.*, 2001].

Tendências gerais como sejam a globalização e a *Internet* tornam extremamente difícil que se possa ter conhecimento sobre forças competitivas.

O aumento da oferta faz com que os consumidores alterem, rapidamente, as suas preferências o que representa, também, um risco pois os ciclos de vida dos produtos tornam-se cada vez mais curtos.

Um exemplo bem representativo desta situação é o mercado dos computadores pessoais.

Há, motivado pelas razões expostas, uma maior competitividade o que conduz ao aumento da importância que a organização tem que dar ao CCC e à necessidade de se preocupar com a gestão deste conhecimento, tendo em atenção a sua evolução.

2.6 Obtenção do CCC

Tradicionalmente, o CCC tem vindo a ser obtido por pesquisa de mercado com todos os problemas que este tipo de pesquisa acarreta, nomeadamente, a necessidade de

efectuar inquéritos em grande escala, o elevado número de horas despendido na sua distribuição, registo e análise dos resultados ou a falta de fiabilidade dos resultados obtidos pois os inquéritos apenas traduzem intenções e opiniões sobre o produto ou serviço que a empresa oferece, não traduzindo, necessariamente, o comportamento dos clientes.

O conhecimento sobre o comportamento dos clientes obtido através das pesquisas de mercado, conduz a campanhas que visam influenciar esse comportamento e na sequência das quais se podem fazer novas pesquisas, tentando obter quer novo conhecimento quer a confirmação do conhecimento já existente. Aquelas campanhas podem tomar, segundo os objectivos a alcançar, a forma de promoções nos centros de venda, envio de *mail* directamente aos clientes e fornecimento de cupões de desconto ou brindes, entre outros.

Estes métodos, embora permitam agrupar os clientes em proveitosos, potencialmente proveitosos ou não proveitosos no sentido de envidar esforços para a sua retenção, correm riscos pois não são suficientemente flexíveis e, por vezes, baseiam-se em pressupostos que podem não ser correctos. Além disso, a sua acção visa uma modificação do comportamento dos clientes no futuro não tendo em atenção, normalmente, como ele se comportou no passado.

Sabe-se que um dos processos mais fiáveis para prever o comportamento futuro no ser humano é o seu comportamento passado; é, por isso, que a informação obtida a partir das transacções dos clientes pode desempenhar uma importância fundamental no negócio.

Ora, a generalidade das empresas possui registos sobre os clientes os quais são obtidos no acto de compra dum bem ou solicitação dum serviço; esses registos podem permitir a obtenção de conhecimento sobre os clientes, na forma de padrões de comportamento. O conhecimento assim obtido é mais fiável do que aquele que se pode inferir a partir das intenções ou opiniões expressas pelos clientes uma vez que se refere a factos objectivos registados em consequência das suas acções.

Como esses registos são, normalmente, em número elevado e estão a ser constantemente actualizados com novas transacções, é impossível a sua manipulação e análise manualmente; assim, as empresas necessitam de meios informáticos para a realização dessas tarefas²⁹.

Esses meios informáticos devem permitir o registo dos dados, a sua manipulação e actualização, uma pesquisa a partir da qual se possam obter padrões de comportamento dos clientes e a actualização e armazenamento destes padrões a fim de serem utilizados em actividades na organização, tendo em vista uma evolução do negócio.

Novos registos podem evidenciar alterações nos padrões de comportamento sendo, portanto, importante que existam mecanismos que, além de criarem novos padrões, permitam, também, a actualização dos já existentes de modo a que fique registada a sua evolução.

As empresas que já estão sensibilizadas para o valor que possuem nas suas bases de dados tentam, com iniciativas diversas, aumentar esse valor. Tal é o caso de numerosas empresas que, cada vez mais frequentemente, estão a fomentar o uso dos cartões de fidelidade para recompensar os clientes mais assíduos. O conceito subjacente é o de que os clientes nestas condições têm tratamento especial tal como descontos específicos nos produtos a adquirir ou em serviços requisitados, numa tentativa de que, para o cliente, seja essa empresa a opção para compras e não a concorrência.

É evidente que a intenção é, mais do que beneficiar o cliente, obter registos que permitam à empresa conduzir uma campanha de mercado individualizada.

Exemplos típicos são o cartão de fidelidade do Jumbo ou o da TAP; outros exemplos semelhantes mas que envolvem parceiros, são o cartão Galp³⁰ ou o ClubSmart da Shell³¹.

²⁹ Os registos aumentam rapidamente quer pelo aumento do número de clientes quer pelo maior detalhe que se considera para cada cliente.

³⁰ www.fastGalp.pt

³¹ www.shell.pt/pt-pt/content/0.6472.37357-61411.00.html

2.7 CCC e Conhecimento Organizacional

Numa organização, normalmente, quando se usa o termo conhecimento [Prusak, 1997] pretende-se dizer experiência, conceitos, valores, crenças e maneiras de trabalhar que podem ser partilhadas e comunicadas.

"**Conhecimento organizacional** é algo de que os agentes organizacionais poderão precisar de saber para executar as suas acções: O que fazer? A quê? Quando? Como? Em que estado se encontra a organização? Qual o estado do seu ambiente? Que transacções foram efectuadas?" [Carvalho & Morais, 2001].

Face a esta definição, é óbvio que o CCC faz parte integrante do CO; para a sua obtenção pode-se considerar dois tipos de conhecimento [Carvalho & Morais, 2001]:

- **conhecimento factual** que é composto por factos que existem e ocorrem resultantes das aquisições de produtos efectuadas pelos clientes ou da solicitação de serviços que a organização disponibiliza;
- **conhecimento conceptual** que é formado por um conjunto de conceitos resultantes da actuação de mecanismos de inferência sobre o conhecimento factual.

Ambos os tipos de conhecimento são registados formalmente mas cada um tem as suas características próprias que o distingue do outro.

A existência de uma classificação de conhecimento organizacional³² ajuda-nos a perceber melhor onde cada tipo de conhecimento fica integrado.

Qualquer que seja a classificação aceite, existem processos tendo em vista a transformação de uns tipos de conhecimento noutros; essa transformação é vantajosa nas

³² Uma revisão de classificações de conhecimento organizacional pode ser encontrada em [Lopes *et al.*, 1999]

empresas, sobretudo quando se pretende conduzir a uma formalização do conhecimento que assim é mais fácil de aceder e gerir.

Uma possível classificação, tendo em atenção a possibilidade de formalização do conhecimento, é a que divide o conhecimento em explícito e tácito sendo este último um conhecimento pessoal, difícil de formalizar ou comunicar a outros enquanto o conhecimento explícito é conhecimento formalizado, facilmente transmissível entre indivíduos e grupos [Nonaka & Takeuchi, 1995].

O conhecimento tácito é o conhecimento que um indivíduo adquire por se ter dedicado a uma actividade durante muito tempo e cuja formalização nem sempre é fácil.

Pode-se incluir, nesta categoria, o conhecimento que os gestores das pequenas empresas possuem, no geral, sobre o comportamento dos seus clientes.

O conhecimento explícito pode ser representado por exemplo, como regras, fórmulas matemáticas ou outras especificações [Choo, 1996].

O conhecimento conceptual é um exemplo de conhecimento explícito; neste conhecimento podemos incluir, por exemplo, o conhecimento que as grandes empresas possuem sobre o comportamento dos seus clientes.

As duas categorias de conhecimento são complementares. O conhecimento tácito enquanto permanece fechado como *know-how* pessoal, é de valor limitado para a organização e tem, entre outros, o perigo do indivíduo detentor do conhecimento abandonar a organização, causando uma perda irreparável para a mesma.

Alguns autores como Choo [Choo, 1996] acrescentam, ainda, um terceiro tipo de conhecimento a que chamam conhecimento cultural e onde incluem:

- as estruturas cognitivas e afectivas que são usadas para perceber, explicar, desenvolver e construir a realidade;
- as hipóteses e crenças usadas para descrever e explicar a realidade;
- as convenções e expectativas que são usadas para atribuir valor e significado à informação.

2.8 O Percurso do CCC

As empresas necessitam de conhecer o comportamento dos seus clientes, conhecimento esse que é obtido a partir dos registos das acções efectuadas por esses clientes e que as empresas possuem nas suas bases de dados.

Na maior parte das organizações o CCC já existe bem como as tecnologias usadas para a sua obtenção e transformação. No entanto, este conhecimento é guardado mas não é gerido, i.e., não é confrontado com anterior conhecimento já existente, consolidando o novo conhecimento, resolvendo potenciais conflitos, actualizando-o e acrescentando, ainda, justificações pertinentes para a evolução temporal verificada; e esta gestão é necessária pois face à evolução dinâmica do conhecimento, acções baseadas no comportamento dos clientes que hoje podem ser muito efectivas já não o serão amanhã.

Uma eficiente gestão do CCC pode contribuir para que a organização tenha vantagem competitiva nos seus mercados de actuação.

E é este novo conceito de gestão do conhecimento conceptual obtido sobre os clientes que já existe em bases de conhecimento criadas a partir da actuação de ferramentas de *Data Mining* (DM) que é o grande objectivo do trabalho presente.

Nas bases de dados estão registados factos; são coisas que existem e ocorrem resultantes das interacções dos clientes com a organização.

Mecanismos de inferência (por exemplo, o MS-Excel ou ferramentas de DM num processo de DCBD), suficientemente explorados e testados, transformam estes factos em conhecimento conceptual.

A Figura 2.4 esquematiza o percurso do CCC.

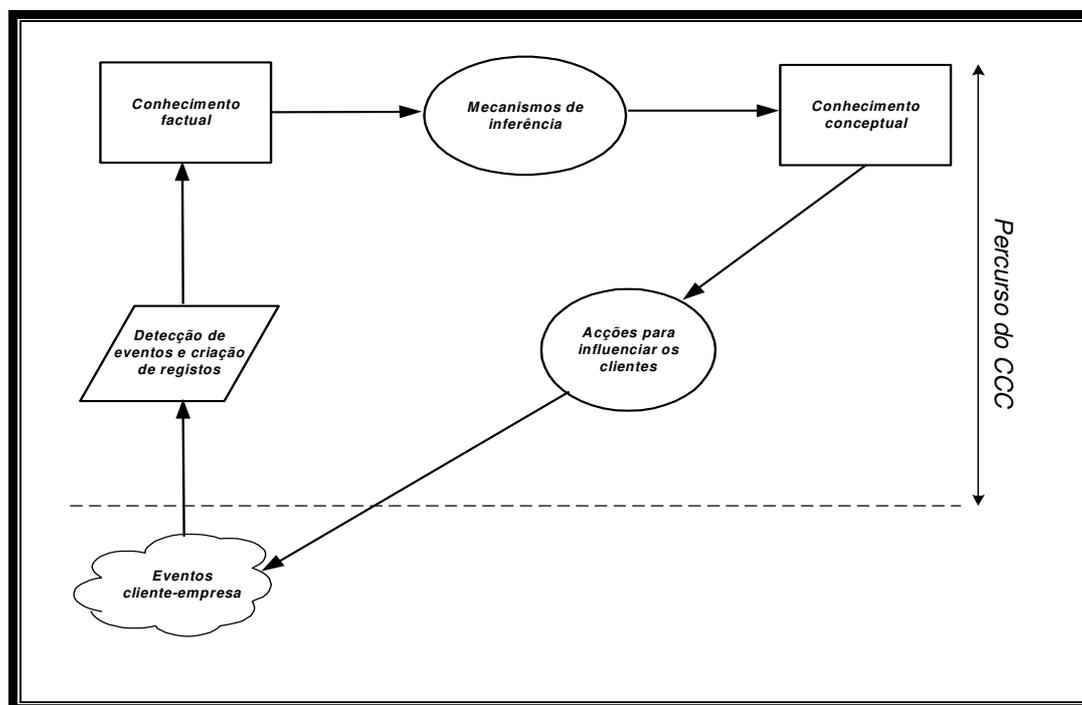


Figura 2.4 - Percurso do CCC.

A análise dos factos conduz a um conhecimento conceptual na forma de [Carvalho & Morais, 2001]:

- descrições;
- classificações;
- tendências;
- relações causa-efeito.

As *descrições* apresentam as propriedades dos factos que são significativas em determinados contextos

Exemplo: A aquisição de uma viagem a Cuba, com uma estadia de 8 dias/9 noites no valor de 900 euros, com partida no dia 1 de Agosto de 2001.

As *classificações* conduzem à distribuição dos dados de acordo com características comuns.

Exemplo: 60% dos agregados familiares com apenas dois elementos e idades compreendidas entre 40 e 50 anos, passam férias na América do Sul.

As *tendências* permitem a detecção de determinado comportamento dos dados.

Exemplo: Da análise dos dados conclui-se que, nos últimos dois anos, a procura de cruzeiros turísticos para os Galápagos aumentou em 25%.

As *relações causa-efeito* são relações entre variáveis em que os dados de uma ou mais variáveis (a causa) são relacionados com os dados de outra ou mais variáveis.

Exemplo: No Verão há uma maior procura de apartamentos turísticos no Algarve.

Os factos registados vão sendo alterados e os mecanismos de inferência criam, em consequência, novo conhecimento conceptual. É, pois, importante que existam mecanismos que, além de se limitarem a criar este novo conhecimento, permitam, também, a actualização do já existente de modo a que fique registada a sua evolução temporal.

A partir daquele conhecimento, são desenvolvidas acções que influenciam o comportamento dos clientes ou criam condições na empresa para a sua satisfação.

Suponhamos, por exemplo, uma operadora da TV Cabo cuja função é atender, telefonicamente clientes que pretendem fazer assinaturas de novos pacotes de canais e que tem, ao mesmo tempo, acesso a um sistema interactivo de padrões de CCC. À medida que o cliente vai respondendo a um questionário no qual são definidas algumas das suas características pessoais, o sistema vai alertando, dinamicamente, a operadora sobre os pacotes a oferecer a um cliente com esse perfil e as perguntas que deve efectuar para melhor explorar a oportunidade de cativar o cliente³³. Esta actuação vai permitir

³³ Adaptado de [Feldens, 2000].

que o cliente fique satisfeito com o modo como foi tratado, tornando-se, desejavelmente, num cliente leal à empresa.

Como se viu, o conhecimento conceptual é criado; no geral, a sua actualização, a sua memorização e o acompanhamento da sua evolução não são feitos.

Deve ser preocupação das organizações cuidar deste conhecimento garantindo que é de qualidade, actualizado e que as pessoas têm acesso a ele.

Tendo em vista esse objectivo, o primeiro passo após a obtenção do CCC é formalizá-lo.

A importância da formalização do conhecimento é fundamental para que:

- haja uma maior certeza de que é utilizado;
- seja mais versátil na sua utilização;
- haja o registo de uma evolução temporal que ultrapassa a capacidade da memorização dum ser humano.

2.9 Sumário

Todo o conhecimento organizacional e, em particular, o CCC, não é estático mas evolui ao longo do tempo e é preciso estar atento a esta evolução e registá-la; o conhecimento desta evolução poderá levar as empresas a repensarem a sua actividade de CRM.

Alterações ao meio ambiente ou acções da própria organização tendo em vista melhorar os serviços prestados ou provocar reacções dos clientes, levam à alteração do comportamento de compra dos clientes. Estas alterações que ficam registadas nas bases de dados através de novas aquisições, mudarão os padrões de compras e, consequentemente, o conhecimento organizacional.

O conhecimento é registado num repositório mas está sempre a mudar; tão depressa obtemos um padrão de comportamento já nos podemos estar a confrontar com a sua

alteração. Há, pois, que gerir aquele conhecimento, tendo em vista a sua actualização, memorização e utilização em acções diversas na organização.

Apresentámos neste capítulo a CRM, uma prática organizacional, essencial para a sobrevivência e desenvolvimento de qualquer empresa na época actual.

Seguir uma prática de CRM pressupõe, em primeiro lugar, ter uma noção exacta de quem é o cliente do negócio e, em seguida, adquirir conhecimento sobre o comportamento desse cliente e geri-lo.

Debruçámo-nos, então, sobre o CCC, a importância que actualmente assume nas organizações e a possibilidade da sua obtenção a partir dos registos dos eventos cliente-empresa.

Este conhecimento efectua um percurso que vai desde a obtenção e registo dos factos - compras de produtos ou solicitação de serviços - passando por mecanismos de inferência que o transformam em conhecimento conceptual que é utilizado, posteriormente, em acções para influenciar a relação cliente-empresa.

Para que este percurso se cumpra, é necessária a existência de tecnologias adequadas que permitam a exploração e gestão de extensos conjuntos de dados.

3. Tecnologias para Aquisição e Registo de Conhecimento

*“The man is an animal who makes tools.”
(Benjamin Franklin³⁴)*

São diversas as tecnologias que dão suporte às actividades que facilitam a obtenção e transformação do conhecimento factual em conceptual, conforme ilustrado na Figura 2.4. Essas tecnologias (Figura 3.1) são enumeradas e descritas neste capítulo.

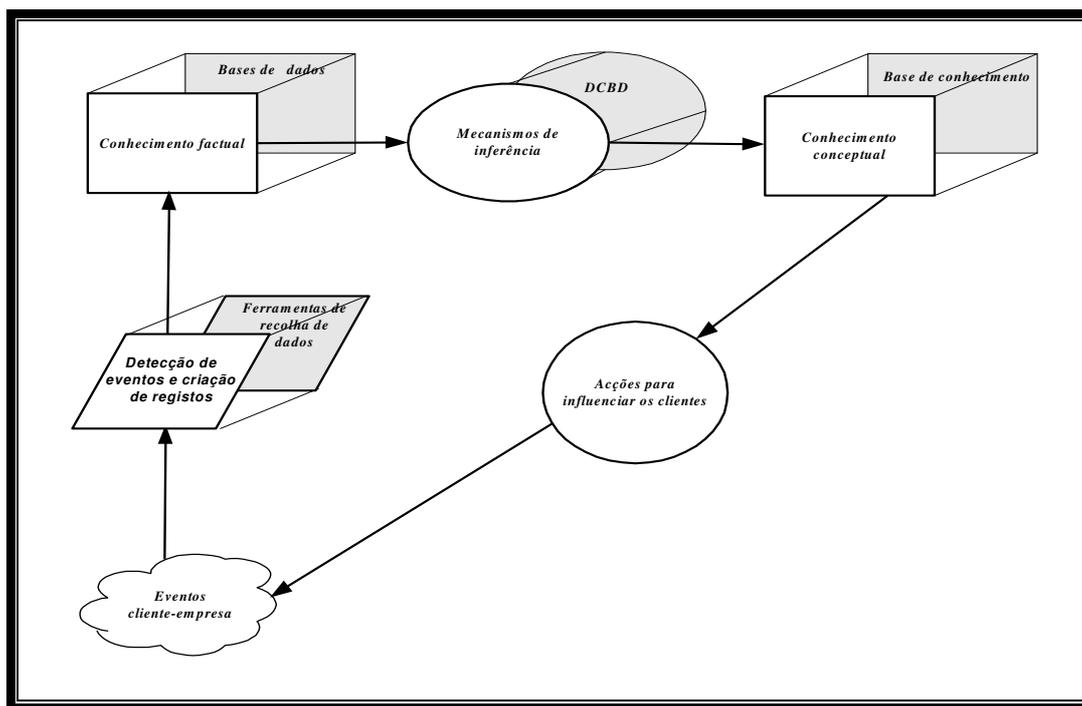


Figura 3.1 - Tecnologias para obtenção de eventos e sua transformação em conhecimento.

O conhecimento factual é normalmente registado em bases de dados através do recurso a tecnologias para a detecção de eventos e criação de registos (e.g., leitores de

códigos de barras, sistemas de reconhecimento de padrões de imagens, sistemas de reconhecimento de voz, *transponders*, cartões magnéticos).

Têm vindo a ser desenvolvidos mecanismos que transformam o conhecimento factual em conhecimento conceptual (e.g., *data mining*), através de processos de **Descoberta de Conhecimento em Bases de Dados (DCBD)** [Frawley *et al.*, 1991], [Brachman & Anand, 1996], [Fayyad *et al.*, 1996a], [Fayyad *et al.*, 1996d]. Estes mecanismos permitem induzir ou extrair novo conhecimento e armazená-lo em bases de conhecimento para posterior utilização.

3.1 Detecção de Eventos e Criação de Registos

Há alguns anos atrás, os registos das transacções dos clientes eram feitos em papel e, posteriormente, transferidos, manualmente, para computadores a fim de serem executadas as operações normais de uma empresa tais como: a facturação, a gestão de *stocks* ou a contabilidade.

Este era um processo muito pouco eficiente pois, além de exigir uma sobrecarga de trabalho da parte dos funcionários da empresa também criava condições para que se verificasse uma percentagem elevada de erros durante a recolha dos dados e até que muitos deles nunca chegassem a ser registados.

O actual desenvolvimento tecnológico veio corrigir essa prática, permitindo, através da utilização de sistemas de identificação electrónica, entretanto desenvolvidos, a detecção automática da ocorrência dos eventos cliente-empresa e a criação dos respectivos registos em computadores.

Existem numerosos sistemas com aquelas possibilidades como, por exemplo:

1 - Leitores de códigos de barras

³⁴ Benjamin Franklin – Cientista americano do séc. XVIII.

No início da década de 80 começou-se a verificar uma mudança em numerosas empresas, nomeadamente em estabelecimentos de venda a retalho com os terminais de venda equipados com um sistema, na altura inovador, que lhes permitia ler, automaticamente, os códigos de barras³⁵ que eram impressos nos produtos directamente ou nas etiquetas que os acompanhavam [Stefix, 1995].

Os leitores de códigos de barras facilitam o registo das transacções efectuadas nos pontos de venda, incluindo o montante comprado de determinado produto, o preço, a data e hora da compra bem como o terminal de pagamento.

2 - Sistemas de reconhecimento de padrões de imagem

Os sistemas de reconhecimento de padrões de imagem registam, por exemplo, o número de pessoas que entram num determinado recinto e as horas a que o fazem.

3 - Sistemas de reconhecimento de voz

Estes sistemas permitem a recolha de dados a partir da emissão de comandos da voz dos clientes quando estes navegam ao longo de menus de aplicações que aceitam aquela tecnologia.

Citemos, como exemplo, o *Universal Messaging* da Unisys, um sistema que permite o acesso a todas as formas de comunicação (e.g., *fax*, *e-mail*, *voice-mail*) a partir de qualquer tipo de equipamento de comunicação (e.g., *Internet*, telefones móveis e fixos, máquinas de *fax*) em qualquer lugar [Baião, 2000].

4 - *Transponders*

A função básica dum *transponder* é a recepção, amplificação e retransmissão para um computador de determinados sinais para os quais foram programados. Podem ser utilizados em domínios diversos como sejam as vias verdes nas auto-estradas, onde permitem que, entre outra informação, seja registada a identificação de cada veículo, as horas a que entrou e saiu da auto-estrada bem como os trajectos percorridos.

5 - Cartões magnéticos

³⁵ Um código de barras associa um único identificador com um produto. O primeiro protocolo utilizado com código de barras foi o EDI (*Electronic Data Interchange*) [Curry & Curry, 2000].

Existem vários tipos de cartões magnéticos, todos eles permitindo o registo de dados.

Temos, por exemplo:

- Os cartões de crédito que permitem registar não só características dos bens com eles adquiridos mas também os processos escolhidos para os pagamentos [Fayyad, 1996], [Berry & Linoff, 2000], [Bits & Pixels, 2001].
- Os cartões multibanco que, usados nas caixas ATM (*Automatic Transaction Machines*), levam ao registo do número do cartão, o terminal utilizado e o tipo de operação efectuada [Brown, 2000].
- Se um cliente utiliza o seu cartão de fidelidade quando se dirige à empresa emissora, esta pode registar, automaticamente, a identificação de compras de produtos ou requisição de serviços feitas pelo cliente [Gutiérrez *et al.*, 2000], [Newell, 2000a].

Em diversos contextos em que a tecnologia é utilizada para diferentes objectivos, também os dados podem ser registados e transformados em conhecimento conceptual. Citemos, como exemplo:

1- Os *call centers*

Os *call centers* permitem que, através de uma simples conversa telefónica, o cliente (ou potencial cliente) forneça elementos sobre os seus hábitos e preferências que são registados manual ou automaticamente, numa base de conhecimento sobre os clientes [Newell, 2000a], [Brown, 2000].

2 - O comércio electrónico

Com o desenvolvimento do comércio electrónico, um sem número de novas possibilidades de registos de factos são agora possíveis; assim, além de dados sobre compras efectuadas, também podem ficar registadas as páginas que cada utilizador visita, quanto tempo gasta em cada uma, o caminho percorrido e se um anúncio desperta, ou não, a sua atenção.

Por exemplo, um *cookie*³⁶, gravado no computador dum cliente, obtém informação detalhada sobre a máquina que está a ser utilizada, o *browser*³⁷ com o qual o cliente acede à página *Web*, e as preferências de compras ou serviços definidos pelo cliente. O *Web server* pode registar informação sobre cada transacção, nomeadamente [Bits & Pixels, 2001], [Ferrão, 2001]:

- a página acedida;
- último *URL* acedido antes da página corrente;
- o *browser* utilizado para acesso à *Web*;
- a data e a hora da visita;
- o tempo da transacção;
- termos de pesquisa utilizados;
- a identificação do cliente;
- eventualmente, palavras chave utilizadas para, através de motores de busca, chegar à página corrente.

3.2 Tecnologias para Registo de Factos

Os sistemas electrónicos utilizados para a recolha de dados a partir dos contactos cliente-empresa, permitem o seu registo automático em suporte informático, normalmente em bases de dados.

3.2.1 As Bases de Dados

As bases de dados são ficheiros de dados onde estes podem ser guardados de uma maneira estruturada, mantendo-se disponíveis para os utilizadores aos diferentes níveis na organização.

³⁶ *Cookie* é um mecanismo que o servidor pode usar para obter informação sobre o cliente.

³⁷ *Browser* é um programa que permite a navegação na *Internet*.

Nas bases de dados, os dados são registados em ficheiros apropriados, independentemente das aplicações que os utilizarão e dum modo transparente ao utilizador. O Sistema de Gestão de Bases de Dados (SGBD) (Figura 3.2) encarrega-se de fazer o *interface* entre as diversas aplicações e os ficheiros de bases de dados o que permite alterações necessárias e/ou oportunas nas estruturas e conteúdos destes sem que haja necessidade de modificar as aplicações. Novas aplicações podem vir a ser desenvolvidas, usando os ficheiros já criados.

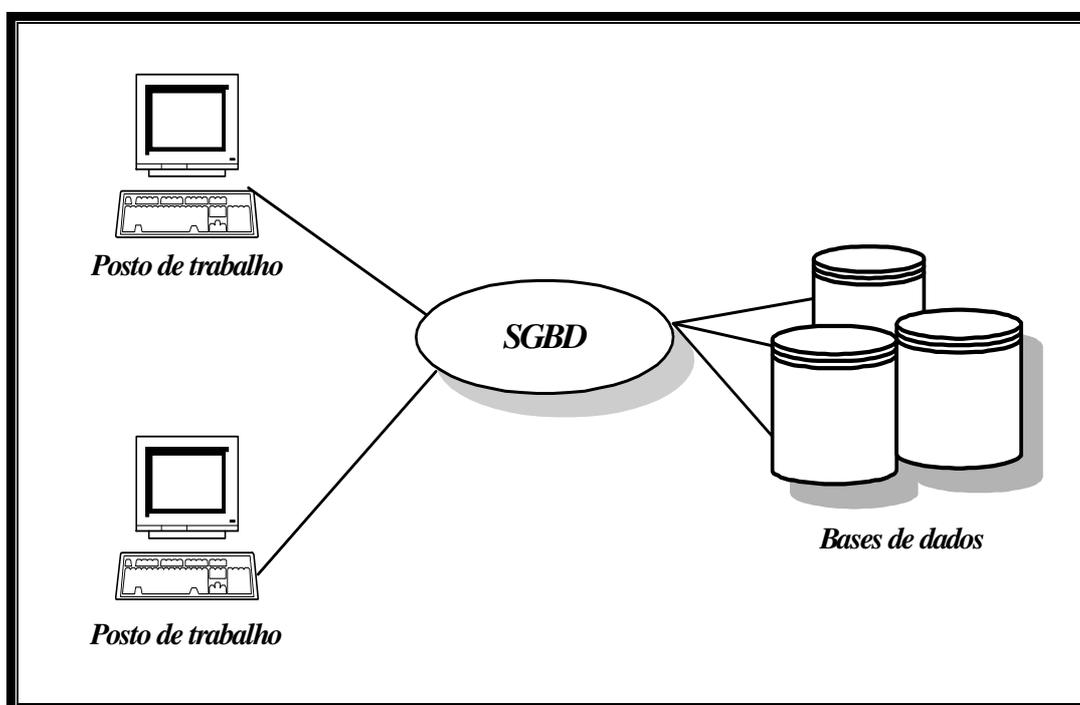


Figura 3.2 - O SGBD.

Vantagens óbvias surgem com a utilização da tecnologia de bases de dados das quais se destacam a redução ou mesmo eliminação da redundância e a independência conseguida entre os programas e os dados [Brathwaite, 1992], [Silberschatz *et al.*, 1996].

Um passo prévio no desenvolvimento de qualquer sistema de bases de dados é a criação do modelo conceptual dos dados [Pereira, 1998], i.e., o desenho de um modelo

que traduza a estrutura lógica dos dados que satisfaz a totalidade dos requisitos da informação dum sistema de informação. Este modelo é um modelo lógico que traduz uma abstracção da realidade desenvolvida a partir de qualquer implementação particular. É claro que esta abstracção vai condicionar os factos que serão registados mas é, até ao presente, o único processo possível.

As bases de dados começaram a aparecer na década de 70 mas tiveram o seu grande incremento no início dos anos 80 com o desenvolvimento do modelo relacional dos dados, dada a flexibilidade da sua utilização que se reduz à realização de operações matemáticas sobre conjuntos.

Atendendo à revolução que sofreu a tecnologia de bases de dados com o aparecimento do modelo relacional, este representa, segundo Date [Date, 1995], a separação entre três gerações distintas de modelos de bases de dados e que são:

- Pré-relacional
 - ▶ modelo hierárquico;
 - ▶ modelo em rede.
- Relacional
 - ▶ modelo relacional.
- Pós-relacional³⁸.

Os modelos da geração pós-relacional, ainda em desenvolvimento, podem-se subdividir em três grupos distintos com características muito próprias e que são [Pereira, 1998]:

- ▶ extensões ao modelo relacional;
- ▶ modelo lógico/dedutivo;
- ▶ modelo orientado aos objectos.

³⁸ Para alguns autores, nomeadamente Stonebraker [Stonebraker *et al.*, 1990], a designação pós-relacional é controversa pelo facto de poder conduzir à ideia de que estes modelos vieram substituir o modelo relacional o que é incorrecto.

Em 1970, Codd [Codd, 1970] lançou os fundamentos teóricos do modelo relacional, estabelecendo toda a base da Teoria da Normalização e um conjunto de regras a que um SGBD deve obedecer para que possa ser considerado relacional.

Pela aplicação da Teoria da Normalização é possível a criação de bases de dados relacionais em que questões como sejam a inconsistência dos dados e a redundância são controladas³⁹.

Toda a Teoria da Normalização se baseia na Álgebra Relacional [Ullman, 1982], [Pereira, 1998] em que o conceito básico é a relação com os respectivos atributos cada um dos quais se define num domínio que não é mais do que o conjunto de valores que esse atributo pode tomar.

Fisicamente, uma **relação** é representada por uma **tabela** em que os atributos são campos (colunas) da tabela, cada um das ocorrências da relação é um registo (linha) da tabela e o elemento na coluna p e na linha x tem o valor $p(x)$ (Figura 3.3).

As diversas tabelas relacionam-se entre si por atributos comuns designados **chaves de ligação**.

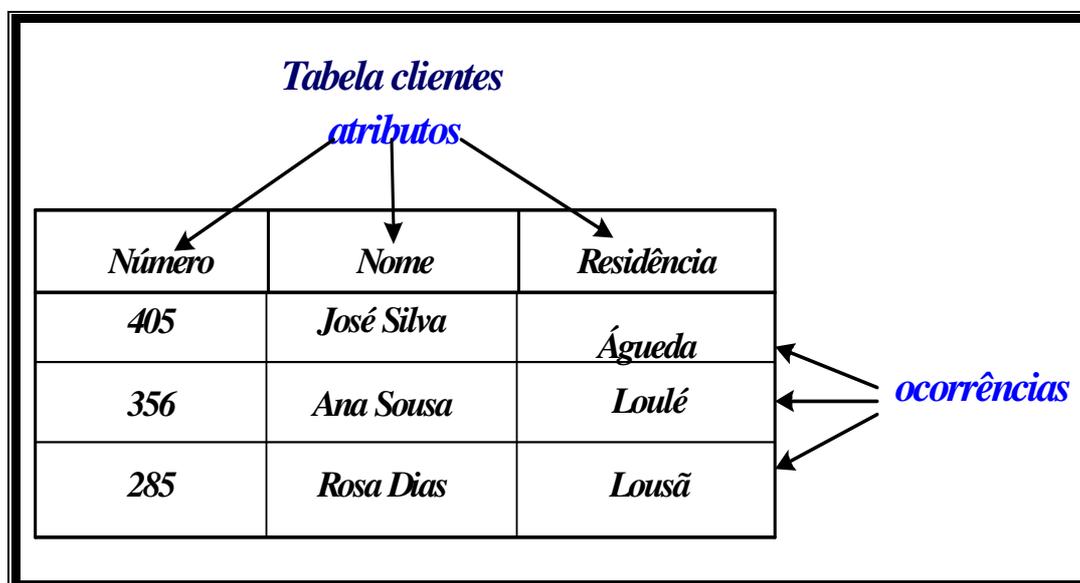


Figura 3.3 – Uma tabela de bases de dados.

³⁹ Em situações específicas, alguma redundância pode ser útil para melhorar a eficiência duma aplicação ou acelerar um determinado processo.

No exemplo seguinte, são apresentados extractos das tabelas clientes (Tabela 3.1), produtos (Tabela 3.2) e transacções (Tabela 3.3) que constituem um sistema de bases de dados de a um supermercado.

Exemplo:

Tabela 3.1 - Tabela de clientes.

n_cliente	nome_cliente	data_nasc	localidade	cod_postal	profissão
127	João Silva	23-05-49	Coimbra	3000	agricultor
145	José Rocha	12-05-74	Oiã	3770	empresário
190	Maria Gomes	03-12-40	Coimbra	3000	médica
230	Elsa Reis	25-07-80	Bustos	3770	estudante
260	Joana Torres	09-11-34	Coimbra	3000	professora
345	Rui Fonseca	01-01-70	Cercal	3770	empresário
450	Rúben Rocha	28-02-54	Vila Verde	3770	agricultor

Tabela 3.2 - Tabela de produtos.

cod_produto	nome_produto	preço_unidade	stock mínimo	stock real
34567	Açúcar	1 euro	800	3500
45623	Arroz	1,25 euros	500	4000
25932	Canetas	0,60 euros	1000	2500
57392	Borrachas	0,75 euros	350	1300
72193	Cadernos	1,2 euros	500	890

Tabela 3.3 - Tabela de transacções.

n_cliente	data	cod_produto	quantidade
450	23-01-2000	34567	5
450	23-01-2000	45623	2
345	31-01-2000	25932	4
260	02-02-2000	57392	20
260	02-02-2000	25932	40
260	12-03-2000	25932	20
260	12-03-2000	34567	5
127	24-03-2000	34567	3
127	24-03-2000	25932	1
230	25-03-2000	57392	1
230	25-03-2000	25932	4
345	25-03-2000	72193	8
145	27-03-2000	25932	12
145	27-03-2000	72193	20
230	29-03-2000	57392	2
230	29-03-2000	72193	4

A aplicação de conceitos da Álgebra Relacional conduz, por aplicação dum conjunto de axiomas [Ullman, 1982], [Silberschatz *et al.*, 1996], ao estabelecimento das formas normais. A passagem de umas formas normais a outras mais elevadas traduz uma diminuição na redundância acompanhada da satisfação dum conjunto de outros requisitos específicos.

O processo da normalização dá-se por completo, usualmente, quando as relações estão na terceira forma normal⁴⁰, havendo, no entanto, formas de grau superior que, em condições específicas são aconselhadas.

Actualmente, novos modelos estão a ser investigados e desenvolvidos – a geração pós-relacional - tendo sobretudo a preocupação de contemplar situações que não são abrangidas pelo modelo relacional como seja o tratamento de estruturas de dados não convencionais (multimédia, imagem, texto).

O modelo relacional continua, no entanto, a dar resposta à generalidade das situações quando a única necessidade é, apenas, o registo de transacções e os processamentos inerentes às aplicações básicas da organização (por exemplo, facturação, salários, *stocks*). É um modelo de fácil utilização, com vastas provas de eficiência em numerosas áreas e com grande implantação nas organizações.

Novos conceitos na área das bases de dados estão a ser introduzidos, tendo em vista uma melhor eficiência da utilização desta tecnologia, nomeadamente as bases de dados distribuídas ou as bases de dados multidimensionais.

As *bases de dados distribuídas* [Date, 1995], são sistemas de bases de dados cujos dados se encontram fisicamente dispersos por várias máquinas ou em suportes físicos

⁴⁰ Uma relação diz-se que está na *3ª forma normal* se todas as dependências funcionais definidas sobre os atributos da relação forem elementares e directas. Uma dependência funcional $X \rightarrow Y$ diz-se *elementar* se não existir nenhuma outra dependência funcional $Z \rightarrow Y$ com Z um subconjunto de X e diz-se *directa* se não houver transitividades [Ullman, 1982].

diferentes na mesma máquina, ligados por meios de comunicação logicamente integrados.

Existem, assim, vários subsistemas das bases de dados em diversos pontos de uma mesma rede em que cada um pode funcionar independentemente ou em coordenação com os outros e todo este processo é transparente ao utilizador que usa os dados como se estivessem localizados no ponto da rede em que ele está a trabalhar.

A regra geral na distribuição é otimizar o processamento local e, para isso, os dados devem ser colocados, preferencialmente, junto de quem mais precisa deles.

A evolução destes sistemas acompanhou a evolução dos sistemas de comunicações; optar por um sistema de bases de dados distribuídas, permite otimizar o tempo de processamento e resposta e fazer uma melhor gestão dos recursos informáticos existentes na organização.

As *bases de dados multidimensionais* [Kenan Technologies, 1995], [Cabibbo & Torlone, 1997], [Furtado, 2000] são sistemas de bases de dados que permitem o registo e manipulação de grupos de dados relacionados, facilitando uma visualização e análise em diferentes perspectivas. Por essa razão, a implantação, numa organização, de uma base de dados multidimensional pode ser encarada como complementar às bases de dados relacionais existentes.

Actualmente, grande incremento está a ser dado a este tipo de bases de dados dada a facilidade da sua utilização pelas ferramentas OLAP (*On-Line Analytical Processing*) [Thierauf, 1997], [OLAP, 1997], [Parsaye, 1997] que requerem rápido acesso aos dados consolidados para obtenção eficiente de informação estratégica para as organizações.

As bases de dados multidimensionais são especialmente vocacionadas para áreas em que os dados são multidimensionais. Citemos, como exemplo, sistemas de informação geográfica, sistemas multimédia e de imagem, bases de dados científicas, bases de dados estatísticas e todas as bases de dados em que a dimensão temporal representa um papel relevante.

O elemento fundamental das bases de dados multidimensionais é o **array multidimensional** em que cada eixo representa uma **dimensão** (uma das perspectivas dos dados) e cada elemento dentro de uma dimensão é chamado uma **posição**.

Na Figura 3.4 pode-se observar um exemplo dum **array** tridimensional.

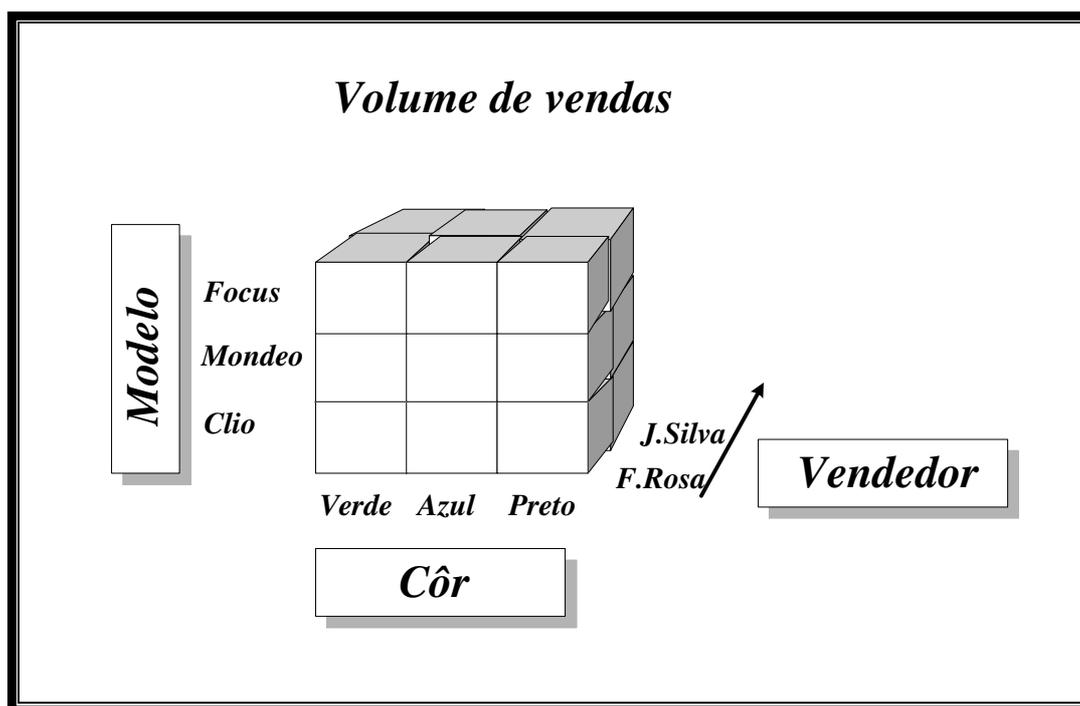


Figura 3.4 – Array tridimensional.

3.2.2. O Data Warehouse

As bases de dados trazem alguns problemas na sua utilização por ferramentas de análise:

1. pelo seu conteúdo (dados frequentemente inconsistentes e campos omissos ou estruturas de dados não uniformes por serem provenientes de aplicações diferentes, instaladas nos mais diversos sistemas):

- ▶ os dados não estão, necessariamente, correctos pois, normalmente, campos como nome, endereço, idade e telefone não passam por uma validação criteriosa nos sistemas transaccionais⁴¹;
 - ▶ existem, frequentemente, definições inconsistentes dos dados e a própria multiplicidade de fontes dificulta a identificação de qual a versão mais correcta, actualizada e adequada; cite-se, por exemplo:
 - ✓ campos em branco como sejam, em ficheiros de clientes, campos de direcção, telefone, data do nascimento;
 - ✓ campos com limites de valores inadequados, como idade de 150 anos ou altura de 5 metros;
 - ✓ codificações diferentes para o mesmo dado, como um código de produto que pode apresentar codificações diferentes em sistemas diferentes;
 - ✓ informações desactualizadas ou divergentes como sejam direcções de clientes em sistemas diferentes;
2. pelo grande peso que resulta da actuação directa das ferramentas de análise, nomeadamente OLAP, prejudicando ou mesmo paralisando a execução das aplicações para as quais aquelas bases de dados foram especificamente construídas (e.g., recolha e actualização de dados, gestão de *stocks*, facturação, gestão de clientes);
 3. pelo horizonte histórico que contemplam ser muito limitado;
 4. pela heterogeneidade dos sistemas operativos sobre os quais estão instaladas.

Por esse motivo, numerosos esforços foram feitos no sentido de encontrar uma solução que satisfizesse as necessidades de consultas complexas, na maioria das vezes com sumariações quase permanentes sobre dados de qualidade garantida e sem prejudicar o normal funcionamento do registo dos dados e execução das operações diárias da organização.

⁴¹ **Sistemas transaccionais** são sistemas que visam a detecção e registo de eventos organizacionais, especialmente eventos que correspondem a transacções entre a organização e os seus clientes ou fornecedores.

Uma primeira solução foi apresentada pela IBM em 1991, utilizando o termo *Information Warehouse* que definiu como “um conjunto de SGBD, *interfaces*, ferramentas e facilidades que gerem e distribuem informações de confiança, oportunas, correctas e compreensíveis sobre o negócio para pessoas autorizadas a tomar decisões” [Cabena *et al.*, 1997] .

O conceito evoluiu e, em 1992, Inmon [Inmon, 1992] apresentou uma nova tecnologia já bem definida e precisa – o *data warehouse* - definindo-a como um repositório de dados e informação para suporte ao processo de tomada de decisão, física e logicamente separado do sistema operacional, cujas características fundamentais são:

1) **Orientado ao assunto** - o *data warehouse* focaliza-se nas entidades com interesse para o negócio (e.g., clientes, vendedores, produtos) enquanto nas bases de dados os dados são orientados para aplicações e funções (e.g., vendas, gestão de *stocks*, facturação).

2) **Integrado** - os dados são armazenados num formato consistente.

Esta é, talvez, a característica mais importante do *data warehouse*.

Dados provenientes de sistemas tradicionais da empresa que pretendem referir-se à mesma entidade aparecem com as mais variadas representações; a sua uniformização é necessária antes de integrarem o *data warehouse*.

A Figura 3.5 exemplifica dois exemplos de integrações possíveis:

- Aplicações informáticas diferentes usam diferentes formas de representar o sexo dos indivíduos como, por exemplo:
 - ▶ *m,f*;
 - ▶ *1,0*;
 - ▶ *x,y*;
 - ▶ *h,m*.

Uma opção de integração é a substituição de todas estas representações pela representação única: *m,f*.

- Aplicações informáticas diferentes usam, como unidades de comprimento, nomeadamente, *cm*, *metros* e *polegadas*.

Uma integração possível é a redução de todas essas medidas a uma única unidade: *cm*.

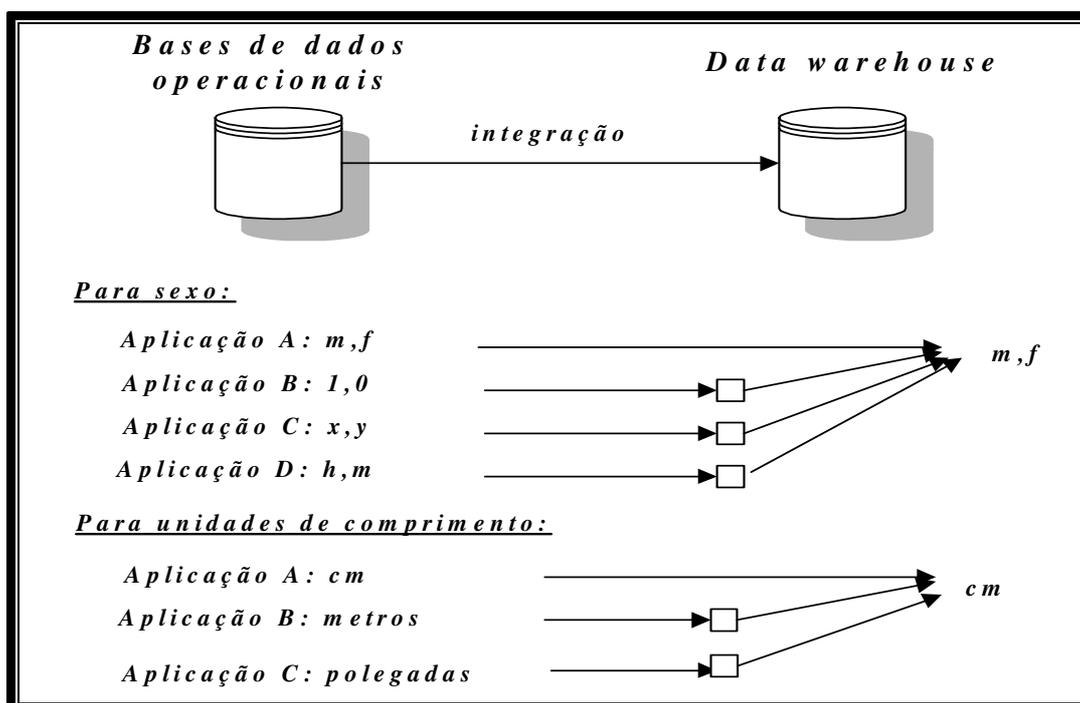


Figura 3.5 – Integração de dados num *data warehouse* (Adaptado de [Inmon & Hackathorn, 1994]).

3) Variante com o tempo – a componente tempo está sempre representada explícita ou implicitamente no *data warehouse*.

Enquanto nas bases de dados os dados reflectem a situação da empresa num determinado instante (e é isso que interessa) no *data warehouse* há um registo dos dados ao longo do tempo, permitindo a análise de uma evolução temporal.

Assim, dum modo geral, as bases de dados contêm registos correspondentes a um período entre 60 e 90 dias enquanto que, no *data warehouse*, os registos existentes correspondem a intervalos temporais de 5 a 10 anos.

Esta característica do *data warehouse* tem, ainda, outra consequência e que é a de que nas bases de dados, os dados estão continuamente a ser alterados e os valores dos dados num determinado instante sobrepõem-se aos valores anteriores enquanto que, no *data warehouse*, novos valores dos dados são acrescentados aos já existentes e com a respectiva componente temporal; não há, aqui, uma modificação de dados existentes.

4) **Não volátil** – uma vez integrados no *data warehouse*, os dados não são alterados nem apagados.

Há, no *data warehouse*, apenas duas espécies de operações, o carregamento regular dos dados e o acesso para consulta, o que leva a que os utilizadores trabalhem num ambiente apenas de leitura.

Pelo exposto, podemos concluir que os dados nas bases de dados e no *data warehouse* se encontram organizados de maneira diferente, tendo em vista o fim a que se destinam.

Pode-se destacar, por exemplo, que:

- os dados são filtrados e sumariados para serem incluídos no *data warehouse* bem como completados com dados provenientes de outras fontes;
- o horizonte temporal dos dados é muito diferente nos dois ambientes;
- os dados nas bases de dados encontram-se todos ao mesmo nível de detalhe enquanto no *data warehouse* há diferentes níveis de sumariação e detalhe conforme se procura ilustrar na Figura 3.6.

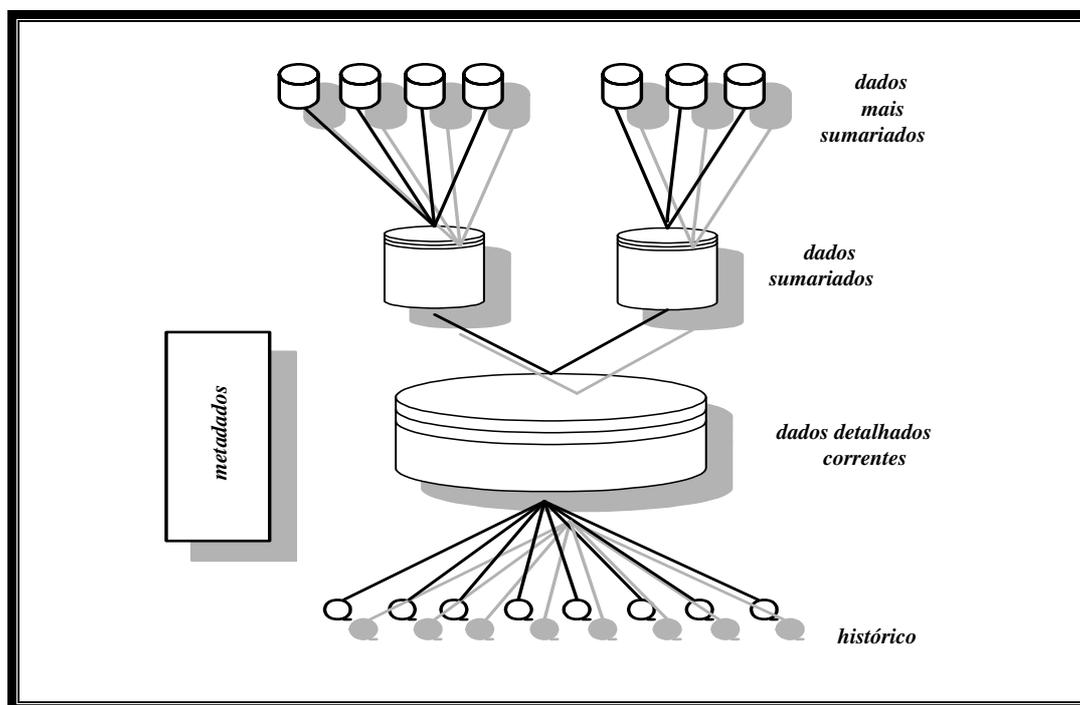


Figura 3.6 – Organização dos dados num *data warehouse* (Adaptado de [Inmon & Hackathorn, 1994]).

Apresentamos, como exemplo, extractos de algumas agregações possíveis entre os dados dum *data warehouse* (clientes com transacções - Tabela 3.4; produtos vendidos por código postal - Tabela 3.5; produtos vendidos por profissão - Tabela 3.6 e produtos vendidos por localidade - Tabela 3.7) obtido a partir do sistema de base de dados exemplificado em 3.2.1.

Exemplo:

Tabela 3.4 - Agregação de clientes com transacções.

data_nasc	profissão	n_cliente	data	nome_produto	quantidade
28-02-54	agricultor	450	23-01-2000	açúcar	5
28-02-54	agricultor	450	23-01-2000	arroz	2
01-01-70	empresário	345	31-01-2000	canetas	4
09-11-34	professora	260	02-02-2000	borrachas	20
09-11-34	professora	260	02-02-2000	canetas	40
09-11-34	professora	260	12-03-2000	canetas	20
09-11-34	professora	260	12-03-2000	açúcar	5
23-05-49	agricultor	127	24-03-2000	açúcar	3
23-05-49	agricultor	127	24-03-2000	canetas	1
25-07-80	estudante	230	25-03-2000	borrachas	1

25-07-80	estudante	230	25-03-2000	canetas	4
01-01-70	empresário	345	25-03-2000	cadernos	8
12-05-74	empresário	145	27-03-2000	canetas	12
12-05-74	empresário	145	27-03-2000	cadernos	20
25-07-80	estudante	230	29-03-2000	borrachas	2
25-07-80	estudante	230	29-03-2000	cadernos	4

Tabela 3.5 - Agregação, por mês, de produtos vendidos por código postal.

mês	nome_produto	cod_postal	quantidade
Janeiro	açúcar	3770	5
Janeiro	arroz	3770	2
Janeiro	canetas	3770	4
Fevereiro	canetas	3000	40
Fevereiro	borrachas	3000	20
Março	canetas	3000	21
Março	canetas	3770	16
Março	açúcar	3000	8
Março	borrachas	3770	3
Março	cadernos	3770	32

Tabela 3.6 - Agregação de produtos vendidos por profissão.

nome_produto	profissão	quantidade
açúcar	agricultor	8
açúcar	professora	5
arroz	agricultor	2
canetas	empresário	16
canetas	professora	60
canetas	agricultor	1
canetas	estudante	4
borrachas	professora	20
borrachas	estudante	3
cadernos	empresário	28
cadernos	estudante	4

Tabela 3.7- Agregação de produtos vendidos por localidade.

nome_produto	localidade	quantidade
açúcar	Vila Verde	5
açúcar	Coimbra	8
arroz	Vila Verde	2
canetas	Cercal	4
canetas	Coimbra	61
canetas	Bustos	4
canetas	Oiã	12
borrachas	Coimbra	20
borrachas	Bustos	3
cadernos	Cercal	8

cadernos	Oiã	20
cadernos	Bustos	4

Em resumo, um *data warehouse* é um sistema composto por:

- um conjunto de estruturas de dados;
- um SGBD⁴² otimizado para atender os requisitos analíticos dos utilizadores;
- aplicações para extrair, limpar, transformar e integrar, no *data warehouse*, os dados provenientes de diferentes origens;
- aplicações responsáveis por disponibilizar aos utilizadores finais acesso ao *data warehouse*;
- um repositório para armazenar e gerir os metadados.

Um *data warehouse* deve conter todos os dados de que os utilizadores necessitam para tomar decisões táticas e estratégicas quer sejam provenientes das bases de dados transaccionais da empresa ou de outras fontes mesmo externas cujo conteúdo completará a informação pretendida.

A importância de possuir dados correctos num *data warehouse* é inquestionável pois os padrões de comportamento obtidos a partir de dados incorrectos não correspondem à realidade e, conseqüentemente, prejudicam acções na organização desenvolvidas a partir da exploração desses padrões.

A correcção e completude [Wand & Wang, 1996] dos dados a fim de atingirem o nível de adequação desejado para a sua integração no *data warehouse* não é uma tarefa simples e pode envolver muito tempo e esforço, inclusive trabalho manual de redigitação de dados, contactos com clientes e o desenvolvimento de ferramentas que automatizem se não todos pelo menos alguns dos processos de correcção e transformação.

⁴² Enquanto que um SGBD orientado para o processamento de transacções tem como principal característica funcionar como suporte de actualizações concorrentes de muitos utilizadores, este SGBD deve ser optimizado para o processamento de consultas complexas e não previstas, servindo-se, por isso,

Pelo exposto se conclui que as fases de aquisição, limpeza e transformação dos dados não podem ser subvalorizadas pois desempenham um papel fundamental no sucesso de futuras utilizações dos dados contidos no *data warehouse*. Além disso, constituem um passo importante para aligeirar o processo de actuação das ferramentas que analisarão os dados.

Algumas características muito específicas de um *data warehouse* são:

- a principal forma de acesso é *ad-hoc*, i.e., não pré-definida nem estruturada;
- a redundância é significativa mas esta característica em vez de ser um problema grave como nas bases de dados é um benefício pois torna mais rápidos os tempos de resposta e como é um sistema apenas de consulta, não se geram inconsistências entre os dados.

Um *data warehouse* é dinâmico por natureza; novos requisitos (funcionais) e novos pedidos aparecem continuamente e, portanto, o sistema tem de estar dimensionado para armazenar novas sumariações de dados e responder a questões onde o tempo de resposta é um factor crítico.

Deste modo, um sistema eficiente quer na sua dimensão quer na velocidade de processamento é fundamental. É, por isso, que, quando se caminha para a implantação deste tipo de repositórios, equipamentos de grandes capacidades de armazenamento e com multiprocessamento são as escolhas mais adequadas [Freitas & Lavington, 1998].

Resumindo tudo o que foi exposto neste parágrafo, pode-se concluir que, sob o ponto de vista da sua utilização, há uma nítida distinção entre as tecnologias de bases de dados e os *data warehouses*, sendo cada uma delas apropriada a determinado tipo de cenário:

de algoritmos "inteligentes" que seleccionam as melhores perspectivas para responder a uma consulta [Parsaye & Chignell, 1993].

- as bases de dados armazenam dados operacionais (normalmente alfanuméricos), utilizados para registar e executar as operações diárias da organização enquanto o *data warehouse* armazena dados integrados, consistentes e históricos, destinados às necessidades da gestão no processo de tomada de decisão (não apenas alfanuméricos);
- dá-se a separação física entre os dois sistemas o que é decorrente das diferenças encontradas nos dados manipulados por cada sistema, no suporte tecnológico necessário, na comunidade de utilizadores e, principalmente, nas características do processamento o que leva a que se estruture o *data warehouse* de uma forma diferenciada, frequentemente multidimensional.
- as aplicações de inserção, actualização e eliminação de dados são efectuadas sobre as bases de dados enquanto que as aplicações utilizadas para analisar o negócio da empresa, ajudando a interpretar o que ocorreu e a decidir sobre estratégias futuras, são efectuadas sobre o *data warehouse*.

É oportuno fazer aqui, também uma referência aos *data marts* [Demarest, 1994], [Berry & Linoff, 1997], [Berson *et al.*, 1999], [Groth, 2000] que podemos definir como "subconjuntos" do *data warehouse*, orientados a um assunto específico.

Os *data marts* têm a vantagem, relativamente aos *data warehouses*, de:

- diminuir, significativamente, os tempos de implementação de uma solução deste tipo, o que representa um benefício em termos de custos para as empresas;
- permitir que as empresas passem a ter soluções mais flexíveis que requerem uma manutenção menor e um melhoramento substancial do seu desempenho.

3.3 Criação e Representação do Conhecimento Conceptual

O conhecimento sobre os factos relacionados com os clientes encontra-se normalmente registado nas bases de dados ou no *data warehouse* da empresa.

Mecanismos de inferência, actuando sobre aqueles repositórios de dados, conduzirão à obtenção de conhecimento de outro tipo, nomeadamente conhecimento sobre o comportamento dos clientes.

Os mecanismos de inferência podem ser humanos, semi-automáticos (por exemplo, o *MS-Excel*, o *Start View* ou o *SPSS*) ou automáticos (por exemplo, DM num processo de DCBD).

3.3.1. A Descoberta de Conhecimento em Bases de Dados

Até ao fim da década de 80, as bases de dados, dadas as tecnologias subjacentes⁴³, eram orientadas quase só para o processamento de transacções, servindo como meros suportes de informação para apoio às actividades correntes da organização.

Os grandes avanços conseguidos na tecnologia de bases de dados ao longo daquela década, aliados ao crescente desenvolvimento tecnológico e à diminuição dos custos de armazenamento, conduziram à informatização dos mais diversos serviços e, conseqüentemente, ao registo das transacções organizacionais.

Vêm-se, pois, as organizações possuidoras dum grande manancial de dados mas sem ferramentas apropriadas para os explorar, conduzindo à descoberta de informação que possa contribuir para uma melhor compreensão do negócio uma vez que proceder a uma observação manual ou utilizar métodos tradicionais de análise de dados (por exemplo, folhas de cálculo ou *queries*) é simplesmente impraticável.

Sem um esforço no desenvolvimento de técnicas de suporte ao processo de descoberta de conhecimento, as organizações acabam perdendo muito do valor dos dados que armazenam.

Começou a haver, então, uma preocupação da comunidade científica para a resolução deste problema.

Surgiu, assim, uma tentativa de, a partir de tecnologia já existente (bases de dados e *data warehouses*), aplicando técnicas e ferramentas conhecidas (técnicas de

⁴³ Os processadores eram lentos e os suportes para armazenamento de dados reduzidos e excessivamente caros.

aprendizagem automática, por exemplo, redes neuronais [Parsaye & Chignell, 1993], [Holsheimer, 1995], [Berson *et al.*, 1999]) e apoiando-se na Estatística, criar um ambiente onde todas estas características actuassem em conjunto a fim de levar novo saber às organizações.

Em 1989, realizou-se, pela primeira vez, um *workshop*⁴⁴ dedicado à DCBD, tendo os seus organizadores, Gregory Piatetsky-Shapiro e William J. Frawley, publicado um livro [Piatetsky & Frawley, 1991] onde reuniram os melhores trabalhos científicos apresentados no referido *workshop*.

A partir desta data, caminhou-se, decididamente, no sentido da DCBD, [Frawley *et al.*, 1991], [Brachman & Anand, 1996], www.Iri.fr/~faye/DataMining/menukdd.html, www.biermans.com/culminating/t2_pages.htm e este termo tem vindo a difundir-se, rapidamente, nomeadamente nas comunidades dos Sistemas de Informação e na da Inteligência Artificial, passando a designar as ferramentas cujo objectivo é encontrar padrões e semelhanças em dados em bruto⁴⁵ [Cabena *et al.*, 1997].

Entende-se por **DCBD** o desenvolvimento e interpretação de padrões descobertos nos dados a fim de decidir quais de entre eles constituem conhecimento de acordo com medidas e limites previamente estabelecidos [Fayyad, 1996].

Como o objectivo da descoberta não está normalmente fixado, os novos padrões revelam, frequentemente, informação até então desconhecida.

Se descrevermos o resultado de DCBD como sendo conhecimento, então o processo de DCBD (Figura 3.7) pode ser encarado como um processo de aprendizagem.

⁴⁴ Este *workshop* realizou-se no âmbito da “*Eleventh International Joint Conference on Artificial Intelligence*”.

⁴⁵ Queremos com este termo referir-nos aos dados que apenas foram registados, não tendo sido submetidos a qualquer tratamento posterior (e.g., limpeza, obtenção de atributos derivados, tratamento de campos omissos); corresponde ao termo inglês *dirty data*.

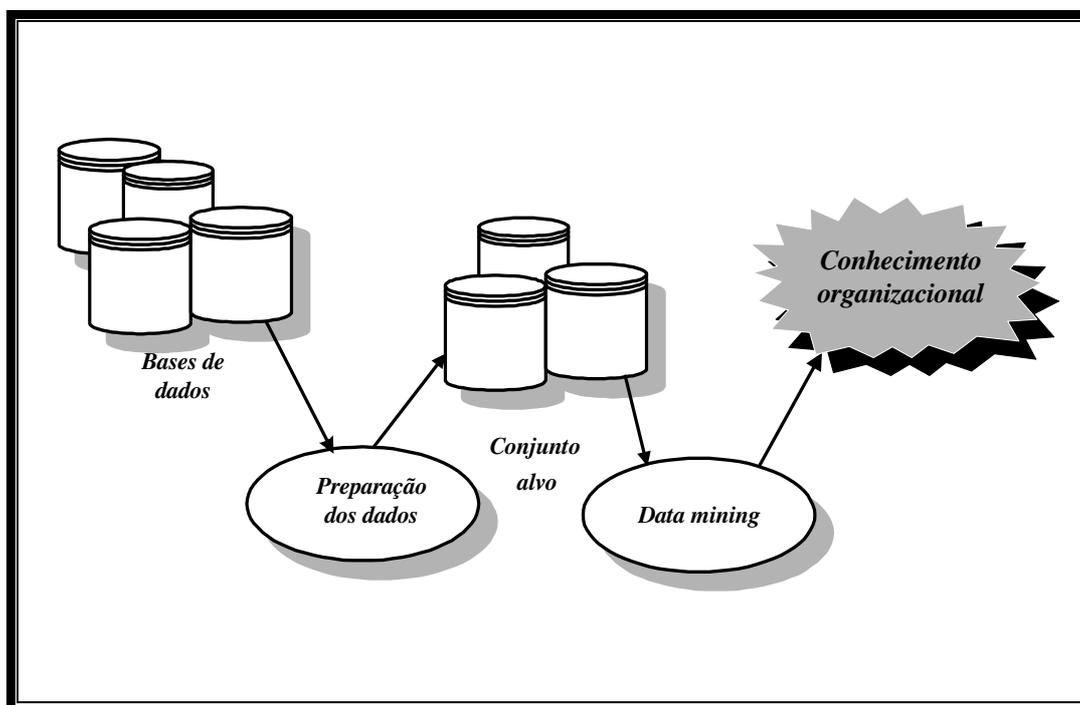


Figura 3.7 – O processo de DCBD.

Os passos principais da DCBD são:

Preparação dos dados - Como as bases de dados são, normalmente, muito grandes, uma pesquisa exaustiva seria muito demorada e, conseqüentemente, muito cara. Para evitar este problema, uma amostra de dados - o conjunto alvo - é seleccionada de acordo com um critério pré-definido⁴⁶; apenas este conjunto é utilizado para a obtenção de padrões. Os resultados obtidos podem, então, ser validados face a toda a base de dados.

Este passo é composto por:

- selecção dos dados;
- pré-processamento;

⁴⁶ É evidente que a selecção dos dados varia com os objectivos do negócio em cada momento [Cabena *et al.*, 1997].

- transformação.

Data mining - Neste passo, o conjunto alvo é pesquisado e algoritmos de aprendizagem automática e estatísticos são aplicados com a intenção de encontrar regularidades⁴⁷, tendências e correlações entre os dados. Os resultados de DM são, normalmente, chamados padrões.

Os padrões descobertos serão, então, consolidados com o conhecimento do negócio⁴⁸ existente⁴⁹. Certamente, este conhecimento existente pode ser usado para seleccionar os dados para formar o conjunto alvo e para planear e executar o passo de DM, sendo utilizado, assim, no contexto do conhecimento prévio. A Figura 3.8 ilustra tal situação.

Entre os benefícios do uso do conhecimento do negócio num processo de DCBD, estão incluídos [Fayyad *et al.*, 1996c]:

- a possibilidade de estabelecimento de um critério para a selecção dos dados para o conjunto alvo e, por isso, a redução do espaço de pesquisa para os algoritmos de DM (ligação 1, Figura 3.8);
- a definição de uma estratégia para que o algoritmo de DM possa efectuar uma pesquisa mais eficiente (ligação 2, Figura 3.8).

As bases de dados estão a ser continuamente actualizadas com novos dados e, por esse motivo, o processo de DCBD deve-se repetir regularmente. Os algoritmos de DM utilizados neste processo, podem usar uma aprendizagem incremental ou não.⁵⁰

⁴⁷ **Regularidade** é uma proposição da forma “ \mathcal{P} verifica-se em \mathcal{R} ”, sendo \mathcal{P} a ocorrência dum determinado padrão e \mathcal{R} o domínio desse padrão. [Klösgen & Zytkow, 1996].

⁴⁸ **Conhecimento do negócio** é todo o conhecimento necessário ao entendimento da forma como o negócio funciona (e.g., o ciclo de vida de um produto, motivos que levam clientes a ter determinados padrões de comportamento na aquisição de produtos, concorrentes).

⁴⁹ Entende-se por **consolidação do conhecimento descoberto**, o processo de incorporação deste conhecimento no conhecimento organizacional ou simplesmente na sua documentação e distribuição pelas partes interessadas; este processo também inclui a procura e solução de potenciais conflitos com conhecimento previamente existente [Fayyad *et al.*, 1996b].

⁵⁰ Uma aprendizagem diz-se **incremental** se, com a actualização da bases de dados com novas ocorrências, o algoritmo de aprendizagem usa a representação do conhecimento já existente; em

Deve-se ter em atenção que o conhecimento do negócio usado na execução de DCBD é o mesmo conhecimento utilizado por outros agentes organizacionais nas suas tarefas operacionais normais ou em tarefas de gestão corrente da organização.

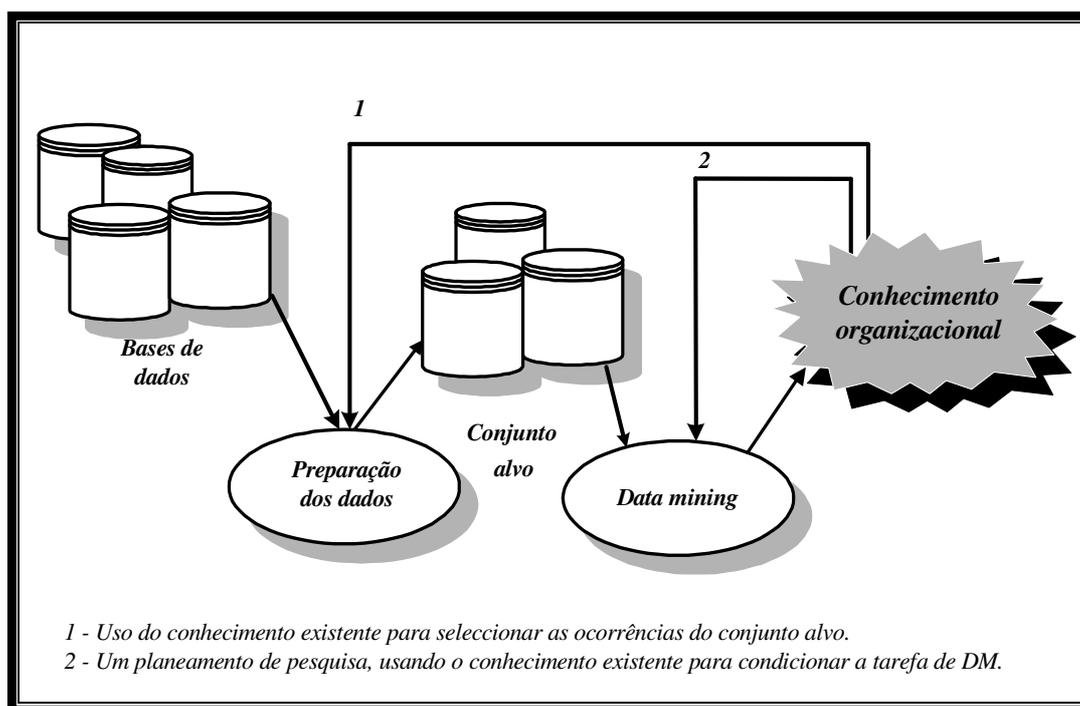


Figura 3.8 - O processo de DCBD usando o conhecimento organizacional.

Durante a fase de recolha de dados, é, por vezes, difícil identificar quais as características que são relevantes e/ou importantes para uma determinada tarefa de aprendizagem e como devem ser representadas. Então, tudo aquilo que se pensa que poderá ser útil, é guardado na base de dados, armazenando-se, por vezes, um sem número de atributos sem interesse imediato. É importante explorar a identificação de quais, entre todos os atributos da base de dados, são realmente importantes para uma determinada operação de DM. Para isso, contribuirá bastante um conhecimento prévio do domínio do problema [Fayyad *et al.*, 1996b].

contrapartida, quando uma nova actuação do algoritmo de aprendizagem não considera a representação do conhecimento já existente, a aprendizagem diz-se **não incremental** [Figuer, 1987].

Este conhecimento contribuirá, por exemplo, para:

- utilização de métodos de transformação ou redução de dimensionalidade [Hu, 1995] a fim de diminuir o número de variáveis em análise e encontrar representações invariantes para os dados;
- conversões de tipos de dados (numérico para caracter ou caracter para numérico);
- utilização de métodos estatísticos para, por exemplo, discretizar um atributo contínuo;
- definição de atributos derivados de outros já existentes, através da aplicação de operadores matemáticos ou lógicos.

Um outro aspecto que deve ser referenciado e que pode melhorar, substancialmente, a eficiência e a qualidade da descoberta de conhecimento em bases de dados é o uso de *data warehouses* como a fonte de dados a ser utilizados na tarefa de DM (Figura 3.9) [Inmon, 1996]. Os registos das bases de dados transaccionais são "limpos"⁵¹ e armazenados em *data warehouses* a partir de onde podem ser usados como fonte de informação para actividades de gestão ou para actividades de investigação no negócio como no caso de DM.

Assim, com o uso de *data warehouses*, problemas como a inconsistência dos dados e o seu armazenamento num formato não consistente são minimizados e os utilizadores de DM podem-se concentrar na sua tarefa sem preocupações com a limpeza e integração.

Com dados não "limpos", algumas variáveis podem possuir valores que sejam significativamente diferentes dos valores esperados e que, em Estatística, são conhecidos por *outliers* [Mello, 1997], [Groth, 2000]; estes *outliers* podem representar novas oportunidades (padrões perfeitamente inesperados) ou podem não ser mais do que dados resultantes de erros variados (por exemplo, uma idade de 300 anos), conduzindo,

⁵¹ *Limpar os dados* significa, por exemplo, corrigir erros de digitação, definir regras para lidar com dados omissos, corrigir notações de restrições de integridade, remover caracteres indesejados, padronizar abreviaturas ou siglas entre outras inconsistências [Brachman *et al.*, 1996].

por isso, a padrões incorrectos [Cabena *et al.*, 1997], [Berson *et al.*, 1999], [Berry & Linoff, 2000].

Exemplo: Foi iniciada uma campanha de venda de fatos de homem em todas as filiais de uma empresa de pronto a vestir da Califórnia. Todas as filiais obtiveram uma grande lucro nas suas vendas à excepção de apenas uma, informação que ficou registada como um *outlier* nos estudos estatísticos levados a cabo pela empresa após a referida campanha.

Uma análise mais cuidada a tal resultado revelou que aquela filial tinha sido a única que optara por dar preferência à rádio, em vez da televisão, na sua campanha promocional.

A detecção de tal situação levou à necessidade da empresa transmitir, às diversas lojas, um conjunto de directrizes no sentido de seleccionarem os meios adaptados à difusão de futuras campanhas.

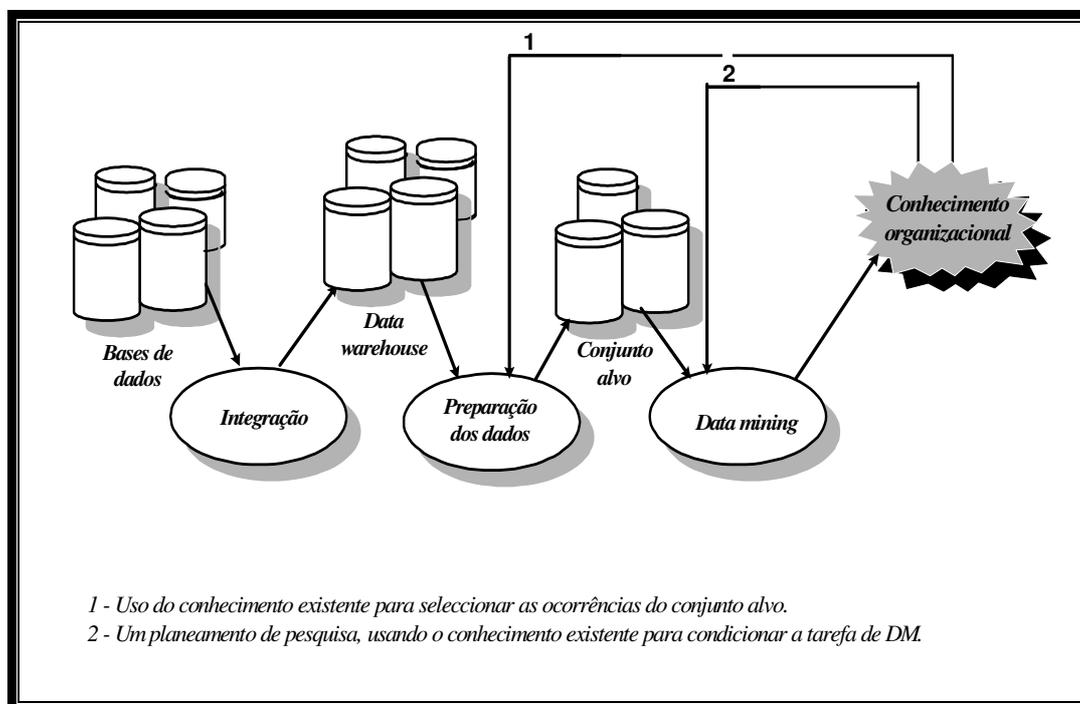


Figura 3.9 - A integração do *data warehouse* no processo de DCBD.

Vemos, pela observação da Figura 3.9 que, para a aquisição do conhecimento conceptual, i.e., para a passagem dos factos a padrões, existem já disponíveis tecnologias que, actuando em conjunto, satisfazem o pretendido e que são:

- as bases de dados;
- os *data warehouses* (ou *data marts*);
- as ferramentas de DM.

3.3.2. As Ferramentas de *Data Mining*

DM⁵² é uma das etapas do processo de descoberta de conhecimento em bases de dados. É, sem dúvida, o elemento mais importante no processo de DCBD e a sua função é obter padrões a partir dos factos registados nas bases de dados ou nos *data warehouses* [Simoudis & Kerber, 1996], [Uthurusamy, 1996], [Han & Kamber, 2001], www.biermans.com/culminating/t2_pages.htm, datamining.br.listbot.com, www.Iri.fr/~faye/DataMining/menukdd.html.

Em 1990, no “*National Science Foundation Workshop*”, Silberschatz, Stonebraker e Ullman, introduziram, pela primeira vez, a expressão *data mining* [Silberschatz *et al.*, 1995], considerando esta como um dos mais promissores temas de pesquisa nas áreas de bases de dados.

A investigação e a exploração nesta área ainda não chegaram a uma evolução tal que se possa estabelecer uma uniformização quer para o conceito quer para o objectivo de DM. Assim, aparecem definições para DM bastante restritivas (“métodos que ...”, “a aplicação de algoritmos ...”) ou excessivamente abrangentes (“qualquer pesquisa manual de dados ...”).

Consideramos que DM é um dos passos do processo de DCBD em que, através de algoritmos oriundos da aprendizagem automática ou construídos especificamente para análise do conteúdo dos dados, são obtidos padrões de comportamento.

⁵² O termo DM deriva da similaridade existente entre pesquisa pormenorizada da informação em grandes bases de dados e a exploração dum filão à procura dum minério com valor.

Independentemente das definições apresentadas, é evidente a necessidade cada vez maior da utilização desta ferramenta como instrumento de exploração do conteúdo das bases de dados, tendo em vista a obtenção de padrões que aumentarão, provavelmente, o conhecimento do negócio.

Numerosas ferramentas foram e estão a ser desenvolvidas nesta área o que prova o grande sucesso conseguido. É difícil, no entanto, ter uma visão deste sucesso na área das aplicações uma vez que os resultados experimentais são, frequentemente, confidenciais dada a sua importância estratégica.

É preocupação, cada vez maior, conhecer o comportamento dos clientes, as suas necessidades individuais e perfis de compras a fim de poder aplicar técnicas de mercado dirigidas a segmentos específicos da população, tendo em vista, por um lado, um acréscimo de lucros no negócio e, por outro, uma oferta de melhores serviços. Assim, há um grande espectro de campos de aplicação onde DM pode ser utilizado: finanças, bancos, vendas a retalho, manufactura, cuidados de saúde e aquisição de dados científicos, entre outros [Holsheimer, 1995], [Simoudis, 1996].

Alguns exemplos de aplicações onde o DM tem sido aplicado com êxito são:

- o sistema *FAIS (Financial Crimes Enforcement Network Artificial Intelligent System)* que identifica transacções financeiras com características que podem indiciar branqueamento de dinheiro;⁵³
- o sistema *Falcon Credit-Risk Assessment* da *HNC Inc.* cuja finalidade é detectar transacções com cartões de crédito que se tornam suspeitas;
- o *Coverstory Spotlight* [Anand & Kahn, 1993], cujo objectivo é analisar o registo das transacções efectuadas em supermercados, usando linguagem natural e

⁵³ Senator, T.E.; Goldberg, H.G.; Wooten, J.; Cottini, M.A.; Khan, A.F.U.; Klinger, C.D.; Llamas, W.M.; Marrone, M.P. and Wrong, P.W.H. - "The Financial Crimes Enforcement Network AI System (FAIS): Identifying potential money laundering from reports of large cash transactions", *AI Magazine*, 16,4, pp.21-39, Winter 1995, apud [Brachman *et al.*, 1996].

gráficos, e fornecendo relatórios sobre as mudanças mais significativas num conjunto de produtos e na sua distribuição;

- o *Time Inc.* é o maior editor de revistas no mundo; nas suas publicações, inclui mais do que 30 títulos, incluindo, por exemplo, as conhecidas revistas *Life* e *Fortune*; utilizando DM, desenvolveu um sistema capaz de, a partir do conteúdo das bases de dados da editora, encontrar causas para o desperdício de papel nas suas máquinas impressoras [Berry & Linoff, 2000];
- O *See5* é uma ferramenta de DM que explora bases de dados de registos ecocardiográficos a fim de identificar perfis de doentes que têm maior probabilidade de contrair determinada doença⁵⁴ [Shindler, 1997].
- o sistema *SKICAT (Sky Image Cataloging and Analysis Tool)*, um projecto da NASA, que é usado para identificar e classificar novos tipos de objectos celestes, usando dados coligidos com instrumentos astronómicos.⁵⁵

Esta não é, ainda, uma tecnologia muito amigável e eficiente como se é levado a acreditar quando se consulta determinada literatura ou se vê o anúncio de comercialização deste tipo de produtos. As ferramentas apresentam uma certa dificuldade de utilização e aquelas que se conseguem manipular com mais facilidade são mais limitadas no seu funcionamento.

Muito importante é ter-se em consideração que a acção das ferramentas de DM termina com o fornecimento de padrões que são visualizados no écran ou registados em ficheiros.

Sobre o interesse que as regras que traduzem padrões de comportamento dos clientes podem ter, vejamos os dois exemplos seguintes [Cabena *et al.*, 1997]:

⁵⁴ O See5 foi aplicado, por exemplo, numa base de dados de registos ecocardiográficos da tiróide a fim de construir regras de diagnóstico relativas a hipotiroidismo, www.rulequest.com/see5-info.html.

Exemplo 1: Uma cadeia de supermercados detectou que nas sextas-feiras, ao fim do dia, indivíduos do sexo masculino, ainda jovens, compravam, com uma elevada frequência, fraldas de papel e embalagens de garrafas de cerveja.

Este é um exemplo dum padrão não intuitivo e do qual a referida cadeia de supermercados pode ter tirado o maior proveito quer pela colocação estratégica dos dois produtos nos supermercados quer pela grande promoção dum dos produtos enquanto o outro poderá ter tido uma subida proporcional.

Exemplo 2: Um revendedor de uma determinada marca de automóveis descobriu, no registo do histórico de compras que possui, que os grandes *sedans* são adquiridos, preferencialmente, por clientes de uma faixa etária mais elevada e ricos o que não lhe terá trazido, concerteza, grande benefício comercial pois os seus competidores todos saberão, provavelmente, a mesma regra.

A descoberta de padrões usando DM, pode, também, conduzir à formulação de novas questões às bases de dados ou ao planeamento de uma nova estratégia para que a ferramenta de DM possa efectuar uma pesquisa mais eficiente.

A investigação em DM, mais do que conduzida por considerações de natureza científica, é-o pelas necessidades dos utilizadores que possuem dados quer em maior número quer mais complexos e diversificados.

Talvez por esse motivo, ainda subsistam muitos problemas não resolvidos nesta área e, conseqüentemente, muitas questões não estão suficientemente exploradas ou permanecem em aberto.

Destaquemos, a título de exemplo:

⁵⁵ Weir, N.; Fayyad, U.M. and Djorgovski, S.G. – “Automated star/galaxy classification for digitized POSS”, *Astron.*, J109.6, 1995, pp 2401-12, apud [Fayyad *et al.*, 1996c].

- a qualidade dos dados [Brachman *et al.*, 1996];
- a conciliação do processo de extracção do conhecimento das bases de dados com a protecção à privacidade dos dados [Fayyad *et al.*, 1996c], [Han & Kamber, 2001];
- até onde é aconselhável ou não a interferência dos utilizadores;
- a selecção, entre as regras obtidas com uma ferramenta de DM, das que são interessantes⁵⁶ [Frawley *et al.*, 1991];
- a integração de DM com outros sistemas (e.g., sistemas de bases de dados, sistemas de *data warehouse*, sistemas de bases de dados na *Web*) [Simoudis, 1996], [Han & Kamber, 2001];
- a integração de DM com técnicas de análises de dados geoespaciais, multimédia e texto [Han & Kamber, 2001];
- o desenvolvimento de ferramentas para visualização e análise dos padrões descobertos [Uthurusamy, 1996];
- a padronização de uma linguagem para DM, melhorando, assim, a interoperabilidade entre diferentes sistemas de DM [Han & Kamber, 2001];
- a determinação de processos para a selecção da ferramenta de DM mais adaptada a determinado domínio [Uthurusamy, 1996].

É, provavelmente, um pouco cedo para ponderar o futuro de DM mas já se visualizam algumas tendências.

O conhecimento do processo de DM já não está restringido às comunidades de investigação, estando a transferir-se, rapidamente, para o campo dos utilizadores, tal como acontece com os *data warehouses*. A fase final da evolução será, provavelmente, a sua integração no processo de BI (*Business Intelligence*) [Brackett, 1999], [Kudyba & Hoptroff, 2001], tornando-se cada vez mais difícil a separação de DM como uma

⁵⁶ Uma regra pode não ser interessante num determinado contexto por:

- se referir a atributos sem interesse (por exemplo, ter ocorrido um número muito restrito de vezes);
- ser redundante, i.e., poder ser obtida a partir de outras regras já existentes;

entidade própria. No futuro, tal como tem acontecido com todas as grandes tecnologias, DM poder-se-á tornar completamente transparente na sua utilização [Cabena *et al.*, 1997].

3.3.3. Modelos para Representação do Conhecimento Obtido com *Data Mining*

A extracção de conhecimento a partir dos dados vai envolver dois processos: identificar padrões interessantes, úteis e válidos, e representá-los de uma maneira concisa, eficaz e de fácil compreensão pelos utilizadores que deles necessitem, sendo, para tal, necessários modelos específicos de representação de conhecimento que devem ser avaliados quanto à sua validade, utilidade e interpretabilidade.

Entre os modelos para representação do conhecimento obtido com DM podemos citar: as redes neuronais, as árvores de decisão e as regras. Enquanto as redes neuronais são representações para uso computacional, as árvores de decisão e as regras são representações para uso computacional e humano.

Uma *rede neuronal* é um modelo que se baseia no modo como o sistema nervoso humano opera; a sua unidade básica é o neurónio e está organizada em camadas, sendo a primeira camada representada pelos dados de entrada e os valores propagados dum neurónio para outro, na camada seguinte. Os valores iniciais vão sendo modificados ao longo das camadas por pesos associados aos ramos ou ligações, sendo os pesos iniciais perfeitamente aleatórios. A última camada corresponde ao valor pretendido e depende quer dos valores iniciais quer dum conjunto de parâmetros intermédios que são calculados ao longo do processo [Parsaye & Chignell, 1993], [Holsheimer, 1995], [Clementine, 1997], [Berson *et al.*, 1999], [Okell, 1999].

Uma *árvore de decisão* (Figura 3.10) é uma estrutura composta por um grupo de nodos e folhas ligados entre si por ramos [Berson *et al.*, 1999], [Okell, 1999].

-
- ser uma regra que traduz algum conhecimento que se considera óbvio como, por exemplo, o aumento de vendas de gelados no Verão.

Os nodos representam os atributos da base de dados, os ramos representam os valores possíveis para esses atributos e as folhas representam as classes.

Uma árvore de decisão é construída por um algoritmo (e.g., CHAID⁵⁷, C5.0⁵⁸) que divide os registos de uma base de dados em classes em que a probabilidade numa folha varia de umas classes para as outras.

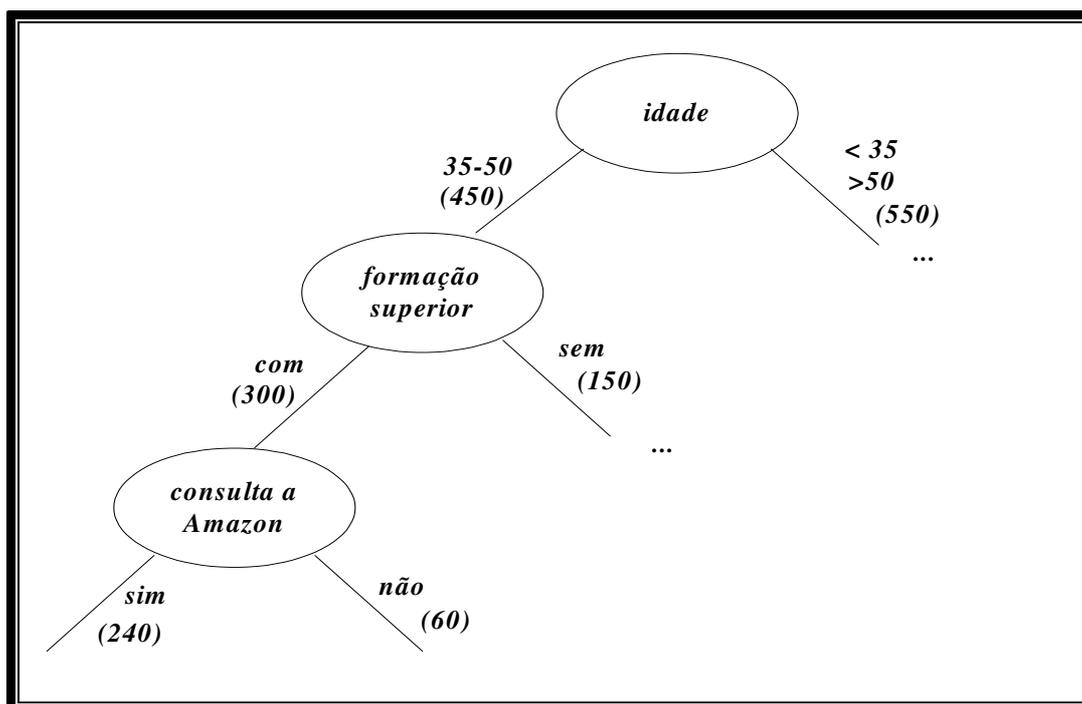


Figura 3.10 – Exemplo de uma árvore de decisão.

Uma **regra** é uma expressão da forma:

$$(3.1) \text{ corpo} \Rightarrow \text{consequente}$$

- *corpo* é um conjunto de descritores;
- *consequente* é um descritor;
- descritor é uma expressão lógica do tipo (atributo, operador, valor) e, como tal, terá sempre associados os valores Verdadeiro ou Falso.

⁵⁷ *Chi Squared Automatic Interaction Detection* [Okell, 1999], [Berry & Linoff, 2000].

⁵⁸ www.rulequest.com, www.cs.umbc.edu/471/hw/hw11/cstutorial.htm.

A expressão (3.1) significa que se o *corpo* tem o valor Verdadeiro então o *consequente* também o tem.

Os métodos que utilizam regras como processo de representação do resultado de DCBD são os mais usados porque mostram grande acessibilidade na manipulação do conhecimento e não apresentam grande dificuldade de compreensão pelos utilizadores da área do negócio. Além disso, todo o sistema resultante fica descrito numa estrutura uniforme o que torna mais fácil a sua compreensão, implementação e manutenção.

Deve-se atender, também, a que a generalidade das situações em que se possa pensar, quer envolva planeamento, diagnóstico, vendas, comportamentos sociais ou outras, pode ser expressa, com facilidade, sob a forma de regras [Dhar & Stein, 1997]; porém, a sua interpretabilidade é discutível quando as regras começam a ser demasiado grandes.

Podemos, usando DM, chegar a três classes de regras distintas:

- sequenciação;
- classificação;
- associação.

R.Agrawal [Agrawal *et al.*, 1993] mostra como as 3 classes de regras apresentadas podem ser encaradas numa perspectiva comum de descoberta de regras o que simplifica bastante o processo de DM.

Regras de sequenciação:

Estas regras referem-se ao ordenamento dos dados e determinam-se, por exemplo, em bases de dados temporais.⁵⁹

⁵⁹ **Bases de dados temporais** são bases de dados que registam informação sobre estados do mundo real através do tempo [Silberschatz *et al.*, 1996].

Dada uma base de dados cujas ocorrências se verificam num certo intervalo de tempo, uma **regra de sequenciação** consiste em encontrar padrões entre as ocorrências em que a presença dum conjunto de itens é seguida por outro conjunto de itens.

Uma análise em regras de sequenciação debruça-se sobre diferentes transações do mesmo cliente ao longo dum certo intervalo de tempo.

Para definir, formalmente, uma regra deste tipo considere-se que $R = \{A_1, \dots, A_m\}$ representa um conjunto de atributos de eventos com os domínios $Dom_{A_1}, \dots, Dom_{A_m}$, sendo um evento e em R um $(m+1)$ -uplo (a_1, \dots, a_m, t) com $a_i \in Dom_{A_i}$ e t um número real que representa o tempo do evento e .

Por uma questão de simplificação, pode-se considerar as seguintes notações:

$e.T$ – o tempo de e ;

$e.A$ – um atributo $A \in R$.

Exemplo: Considerando os registos de acessos a páginas WWW como eventos, pode-se considerar os atributos página (o endereço da página acedida), *host* (o *host*⁶⁰ que acedeu) e tempo (o instante em que a página foi acedida).

Um evento possível : $x.página = www.altavista.com$.

Uma **sequência de eventos** U é um conjunto de eventos definidos sobre o conjunto R , i.e., uma relação sobre $R \cup \{T\}$ em que o domínio do atributo T é o conjunto dos números reais.

Um **episódio** é uma combinação de eventos com uma ordem parcialmente especificada; ocorre numa sequência se há ocorrências dos eventos numa ordem consistente com a ordem dada, dentro dum limite de tempo dado.

⁶⁰ *Host* é o computador ao qual os utilizadores remotos estão ligados [Franks, 1997].

Formalmente, um episódio P nas variáveis $\{x_1, \dots, x_k\}$ que se pode representar por $P(x_1, \dots, x_k)$, é uma conjunção $\bigwedge_{i=1}^k \psi_i(y_i, z_i)$ em que $(y_i, z_i) \in \{x_1, \dots, x_k\}$ são variáveis de eventos e cada conjunto $\psi_i(y_i, z_i)$ tem uma das formas $\alpha(x.A)$, $\beta(x.A, y.B)$ ou $x.T \leq y.T$, em que A e B são atributos de eventos; α e β são funções definidas, respectivamente, por:

$$\alpha: Dom_A \rightarrow \{V, F\}$$

$$\beta: Dom_A \times Dom_B \rightarrow \{V, F\}$$

e x representa o produto cartesiano de dois conjuntos.

Exemplo: Considerando o exemplo anterior, um episódio pode ser:

$$x.página=www.altavista.com \wedge y.página=www.yahoo.com \wedge x.host=y.host$$

Uma **regra de sequenciação** é uma regra do tipo:

$$(3.2) \quad Q_1[V] \Rightarrow Q_2[W]$$

em que Q_1 e Q_2 correspondem a episódios e V e W a números reais.

A regra mostra que se o episódio Q_1 tem uma ocorrência no intervalo $[t, t']$, com $t' - t \leq V$, então o episódio Q_2 ocorre no intervalo $[t, t']$, com $t'' - t \leq W$.

Exemplo:

$$\begin{aligned} & se \ x.página=www.altavista.com \wedge y.página=www.yahoo.com \wedge x.host=y.host \ [60] \\ & \Rightarrow \ x.página=www.altavista.com \wedge y.página=www.yahoo.com \wedge z.página=www.amazon.com \wedge \\ & \ x.host=y.host \wedge y.host=z.host \ [120] \end{aligned}$$

A regra exemplificada dá a indicação de que se algum "internauta" acede às páginas www.altavista.com e www.yahoo.com num determinado instante de tempo, por exemplo, um minuto, é provável que também aceda à página www.amazon.com em dois minutos. [Mannila & Toivonen, 1996]

Regras de classificação:

As ferramentas de DM, actuando sobre os dados, permitem agrupá-los em classes de acordo com semelhanças encontradas e, em seguida, definem as características de cada classe.

Formalmente, uma regra de classificação possui a forma [Holsheimer & Kersten, 1994]:

$$(3.3) \quad \textit{se } D \textit{ então } C$$

i.e., qualquer objecto que satisfaça a descrição D pertence à classe C , em que:

- uma **descrição** D é uma conjunção não vazia de descrições elementares;
- uma **descrição elementar** é uma fórmula $X_1=a_1, X_2=a_2, \dots, X_n=a_n$ tal que:
 1. $X_i \in X$, $i=1, \dots, n$ com X o conjunto de atributos de predição;
 2. $a_i \in \text{Dom}_{X_i}$, $i=1, \dots, n$;
 3. $X_i \neq X_j$ para $i \neq j$, $i, j=1, \dots, n$;
- uma **classe** C_i é um subconjunto do conjunto alvo S formado por todos os objectos que satisfazem à condição de classe (\textit{cond}_i), i.e., $C_i = \{\theta \in S \mid \textit{cond}_i(\theta)\}$, $i=1, \dots, n$.

Exemplo:

Sejam

X_1 o atributo idade e X_2 o atributo formação com $\text{Dom}_{X_1} = \{[5,10],[10,15],[15,20],$

$[20,35],[35,50], \dots\}$ e $\text{Dom}_{X_2} = \{\textit{elementar}, \textit{secundário}, \textit{superior}\}$;

$a_1 = [35,50]$ e $a_2 = \textit{superior}$

a descrição D : $\textit{idade} = [35,50] \wedge \textit{formação} = \textit{superior}$;

a classe C : o conjunto dos indivíduos que regularmente consultam a página www.amazon.com

Formalmente,

$$\textit{se } \textit{idade} = [35,50] \wedge \textit{formação} = \textit{superior} \textit{ então } \textit{consulta a página} = \textit{www.amazon.com}$$

O exemplo apresentado transmite a ideia de que os indivíduos com idades compreendidas entre 35 e 50 anos, com formação superior, consultam regularmente a página www.amazon.com.

Regras de associação:

As ferramentas de DM que geram regras de associação definem associações entre conjuntos de itens com certas afinidades.

Uma análise baseada em regras de associação identifica, por exemplo, itens frequentemente vendidos nas mesmas transacções.

Formalmente:

$$(3.4) \text{ se } X_1=a_1, X_2=a_2, \dots, X_n=a_n \text{ então } Y=b [n_1, n_2]$$

em que:

- $X = \{X_1, X_2, \dots, X_n\}$ é um conjunto de atributos de domínios, respectivamente, $Dom_{X_1}, Dom_{X_2}, \dots, Dom_{X_n}$.
- a_i é uma ocorrência de Dom_{X_i} , $i=1, \dots, n$.
- $X_1=a_1, X_2=a_2, \dots, X_n=a_n$ formam o conjunto dos antecedentes da regra, i.e., o seu corpo;
- $Y=b$ é o conseqüente da regra;
- n_1 , o nível de suporte da regra, representa a percentagem dos elementos do conjunto alvo que verificam $X_1=a_1, X_2=a_2, \dots, X_n=a_n$;
- n_2 , o nível de confiança da regra, representa a percentagem de n_1 que também verifica $Y=b$ [Simoudis, 1996].

O uso de níveis de suporte e de confiança constitui um processo de ajuda para eliminar regras não interessantes, através da definição de limites inferiores para estes parâmetros [Agrawal *et al.*, 1993].

Exemplo:

Sejam

X_1, X_2 e X_3 atributos com X_1 =atividade, X_2 =aceder a página WWW e X_3 = aceder a página WWW;

Y =aceder a página WWW;

$\text{Dom}_{X_1} = \{\text{investigação, docência, profissão liberal, ...}\};$

$\text{Dom}_{X_2} = \text{Dom}_{X_3} = \{\text{www.altavista.com, www.yahoo.com, www.amazon.com, ...}\};$

a_1 =investigação;

a_2 =www.yahoo.com;

a_3 =www.altavista.com;

b =www.amazon.com.

Formalmente,

se atividade=investigação e página de acesso=www.altavista.com e página de acesso=www.yahoo.com

então página de acesso=www.amazon.com [40,70]

O exemplo apresentado mostra que dos indivíduos que acedem à Internet, 40% são investigadores que acedem às páginas *www.altavista.com* e *www.yahoo.com*; destes, 70% também acedem à página *www.amazon.com*.

Algumas das representações de conhecimento definidas podem-se transformar, com facilidade, noutras como seja o caso de uma árvore de decisão em que cada caminho da árvore se transforma numa regra seguindo os princípios:

- os nodos internos e os ramos que deles partem são convertidos em condições do antecedente da regra (parte **SE**);
- o nodo folha é convertido no conseqüente da regra (parte **ENTÃO**).

Exemplo: Dos "internautas" com idade compreendida entre 35 e 50 anos e formação superior, 80% consultam o *Amazon* (Figura 3.10).

Uma regra também pode ser obtida pela associação de vários ramos que conduzem a folhas idênticas.

3.4 O Armazenamento do Conhecimento Conceptual

Normalmente, o conhecimento conceptual, independentemente de qual seja o formalismo utilizado para a sua representação, é guardado em sistemas denominados bases de conhecimento.

O objectivo de uma base de conhecimento [Vogel, 1996] é, pois, registar conhecimento e disponibilizá-lo aos utilizadores que dele necessitem. Deste modo, uma base de conhecimento pode-se comparar a um livro ou a um manual em que informação relevante foi registada tendo em vista disponibilizá-la sempre que necessário.

Frequentemente, as bases de conhecimento não são encaradas da mesma maneira por diferentes comunidades ou contextos.

Neste trabalho, quando usamos esta denominação – base de conhecimento – estamos a referir-nos ao registo ou apresentação de um qualquer formalismo que tem em vista a representação de padrões de comportamento dos clientes num suporte informático.

As ferramentas de DM, pelo facto de criarem conhecimento conceptual, vão alimentar bases de conhecimento cujo conteúdo pode ser expresso segundo diversos formalismos, podendo ir desde um simples relatório ou um gráfico, registados em papel ou visualizados no écran, até a um ficheiro de texto onde as regras são registadas.

Normalmente, estas bases de conhecimento são estáveis pois os padrões gerados pela ferramenta de DM não são objecto de qualquer tratamento mas apenas nelas registados, não existindo um processo dinâmico de actualização e análise da evolução do seu conteúdo; os sistemas existentes limitam-se a acrescentar, na base de conhecimento, a nova informação àquela que já está registada ou, frequentemente, sobrepõem a nova informação à anterior.

É evidente que uma base de conhecimento guardada num suporte informático é muito mais flexível do que se for apenas impressa ou visualizada num écran, podendo servir de suporte activo à resolução de várias tarefas ou ser integrada, dinamicamente,

no conhecimento organizacional da empresa; tem, ainda, a vantagem de permitir uma actualização dinâmica do seu conteúdo.

Para registar uma base de conhecimento num suporte informático, podem ser utilizadas diferentes formalismos de representação de conhecimento (e.g., regras de produção, redes semânticas, enquadramentos).

3.5 Sumário

Neste capítulo foi apresentado um conjunto de tecnologias que permitem a recolha dos eventos cliente-empresa, o seu registo e a sua transformação em padrões de comportamento que também são registados.

A Figura 3.1 ilustra a relação entre as operações de obtenção de eventos cliente-empresa e a sua transformação em conhecimento e as tecnologias que permitem a sua realização.

No entanto, estas tecnologias apenas permitem a realização das operações apresentadas até ao registo do conhecimento, não prevendo a gestão deste conhecimento, no sentido do seu confronto com o conhecimento já existente e sua consolidação.

4. Gestão do Conhecimento do Comportamento dos Clientes

*"We are drowning in information but
starving for knowledge"
(John Naisbett⁶¹)*

No capítulo dois foram apresentados conceitos relacionados com CRM e a sua importância, foi dada particular importância ao conhecimento sobre o negócio (e especialmente sobre o comportamento dos clientes) necessário para a concretização de CRM. Por fim, no capítulo três, foi apresentada a tecnologia disponível que permite a obtenção do conhecimento factual e a criação do conhecimento conceptual.

Neste capítulo, analisaremos, em particular, o CCC, mostrando a utilidade da introdução de um novo conceito: a "**Gestão do CCC**".

Tal como qualquer comportamento humano, o comportamento dos consumidores evolui devido a diversos factores, nomeadamente:

- novas tecnologias que conduzem ao aparecimento de novos produtos;
- mais acesso a informação sobre produtos;
- globalização dos mercados;
- influência de associações de consumidores;
- *marketing* cada vez mais activo e agressivo.

⁶¹ Megatrends - New York: Warner Books, Inc, 1982 apud www.biermans.com/culminating/t2_pages.htm e www.Iri.fr/~faye/DataMining/menukdd.html.

Os padrões de comportamento das pessoas evoluem ao longo do tempo; assim, parece ser razoável que esta evolução do comportamento dos clientes seja acompanhada. Este acompanhamento trará vantagens para a organização como sejam:

- manter actualizado o conhecimento sobre o comportamento dos clientes;
- facilitar a tarefa de compreender e explicar a evolução do comportamento dos clientes;
- conduzir a uma gestão mais eficiente do conhecimento organizacional sobretudo na vertente do conhecimento do negócio.

Um dos objectivos deste trabalho é apresentar uma proposta para a actividade do acompanhamento referido através da introdução de um novo conceito na organização a **“Gestão do CCC”**.

Conceptualmente, a solução aqui proposta, satisfazendo um conjunto de requisitos funcionais (e.g., modularidade, flexibilidade), tem em vista zelar pelo CCC:

- criando e mantendo uma base de conhecimento com os padrões obtidos;
- para cada novo padrão descoberto:
 - ▶ verificar se já era conhecido;
 - ▶ verificar se traduz a evolução de algum padrão já conhecido e registar essa evolução;
 - ▶ consolidar o novo conhecimento, i.e., resolver problemas de interpretação, integração e contradições com padrões já existentes;
 - ▶ armazená-lo;
 - ▶ torná-lo acessível a quem dele necessite;
- anotando a utilidade dos padrões e os resultados da sua utilização em acções organizacionais.

4.1 Modelo Global do Processo de Aquisição, Registo e Exploração do CCC

Dos capítulos 2 e 3, pode-se concluir que a tecnologia existente se limita a registar (ou a apresentar), numa base de conhecimento, o conhecimento obtido sob a forma de padrões de comportamento sem lhe dar qualquer tratamento.

Mas, algumas questões se podem pôr relativamente aos padrões assim registados, como sejam:

- O que se faz, habitualmente, com os padrões obtidos?
- A quem interessam estes padrões?
- Já eram conhecidos na altura da sua obtenção?
- Que benefício tira a empresa da sua criação?
- Constituirão, apenas, mais uma informação a acrescentar às muitas já dispersas pela empresa?
- Que fazer para que não sejam esquecidos?
- Os padrões de comportamento evoluem no tempo. Que fazer para acompanhar essa evolução?

É importante ter presente, ainda, que o facto das bases de dados transaccionais estarem a ser continuamente actualizadas, permite uma análise sobre a evolução do comportamento dos clientes. Gera-se, assim, a necessidade de ferramentas que captem e registem estas mudanças ou não será possível uma análise da evolução temporal dum padrão de comportamento. Esta variação traduz-se pela alteração dos valores dos parâmetros (níveis de suporte e de confiança) da regra que o representa, permitindo a consideração das seguintes possibilidades:

- um crescimento é, provavelmente, indicativo do aumento de interesse, para o negócio, do padrão de comportamento associado à regra;

- uma diminuição progressiva, é sinal de que o comportamento representado por esse padrão está a desaparecer; ao fim de algum tempo dos parâmetros se manterem abaixo dum limite definido, poderá ser considerado o seu abandono;
- alterações periódicas podem ser indicativo de várias situações, nomeadamente, comportamentos sazonais;
- um valor extremo accidental (máximo ou mínimo) significa, concerteza, uma situação anormal cujas razões poderão merecer uma análise cuidada.

O registo dos padrões de comportamento dos clientes e da sua evolução temporal poderá conduzir a dois tipos de acções: acções na organização e acções para influenciar o comportamento dos clientes.

As acções na organização são acções que têm em vista a criação da melhoria de condições para os clientes ou oferta de novos serviços, sobretudo em situações onde é importante antecipar o comportamento dos clientes a fim de que a organização se encontre preparada para lidar com eles.

Cite-se, como exemplos:

- a diminuição nas filas de pagamento nos supermercados;
- o aumento das entradas e saídas em certas auto-estradas nas horas em que se prevê haver maior afluxo de trânsito;
- a apresentação de páginas *Web* com uma informação mais de acordo com os interesses demonstrados pelos utilizadores;
- o encaminhamento automático das chamadas nos *call centers*, sem a passagem por atendedores intermédios;
- a disponibilização de opções variadas na utilização de caixas ATM.

As acções para influenciar o comportamento dos clientes são acções que têm em vista provocar reacções da parte destes, podendo conduzir a uma alteração dos perfis de compras; podem ser, por exemplo, campanhas de *marketing* orientadas para segmentos específicos da população ou mesmo *marketing* individualizado.

Pelo já exposto, pode-se concluir que o modelo para a criação do conhecimento conceptual a partir do conhecimento existente sobre o comportamento de compras dos clientes e sua integração no conhecimento organizacional, está, actualmente, na situação que se apresenta na Figura 4.1.

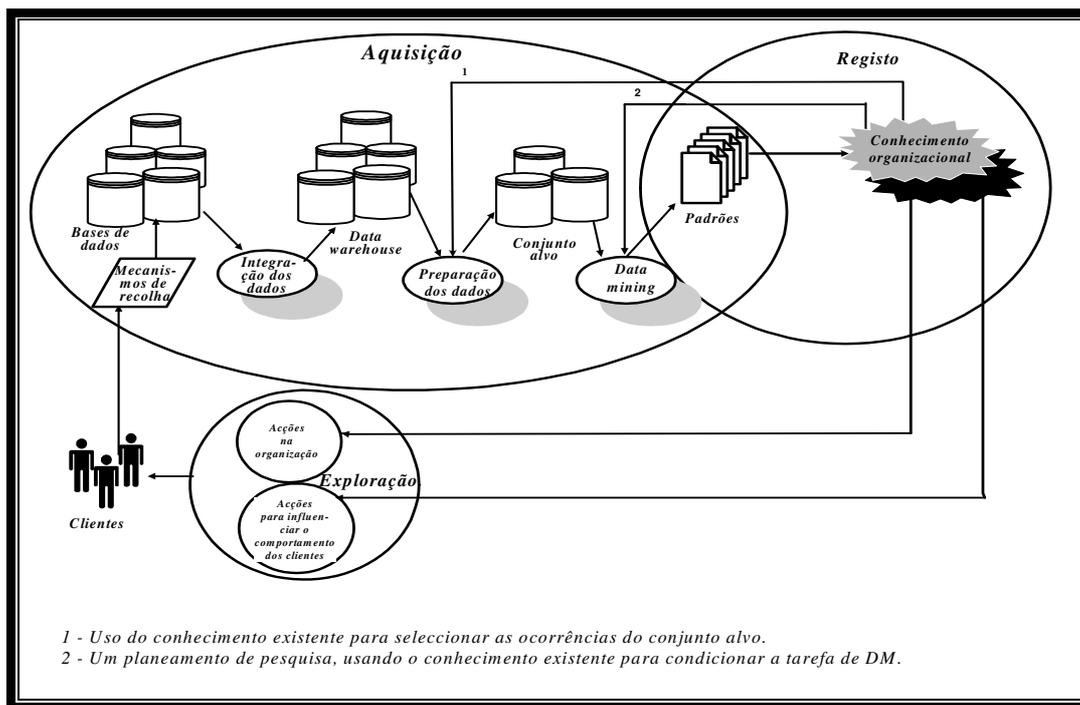


Figura 4.1 - Modelo global do processo de aquisição, registo e exploração do CCC.

4.2 O Cenário do Tratamento dos Padrões

Numa fase de aquisição do CCC, as técnicas de DM permitem automatizar a detecção de padrões a partir das bases de dados transaccionais, através dum processo de DCBD.

No geral [Kohavi & Provost, 2000], os padrões obtidos com a aplicação destas ferramentas são transmitidos, pelo perito de DM, aos agentes organizacionais que têm

os seus próprios processos (manuais ou computacionais) de retenção dessa informação e posterior aplicação por exemplo em campanhas de *marketing*⁶².

Assim, questões importantes como sejam a razão do aparecimento dos referidos padrões e análise da sua evolução não são, usualmente, tidos em consideração.

É frequente que o conhecimento que esses padrões possam fornecer, esteja apenas na mente dos gestores do negócio os quais podem ser deslocados para outros serviços, reformados ou mesmo abandonar a organização, fazendo com que aquele conhecimento se perca⁶³; além disso, é usual esse conhecimento estar disperso pela organização em diferentes formatos ou mesmo não formalizado.

Estes procedimentos têm custos elevados para a organização, nomeadamente de ordem temporal.

Consequentemente, aquele conhecimento poderá não ter grande interesse se não for formalizado, actualizado, validado e integrado, dinamicamente, no conhecimento organizacional a partir de onde pode ser usado em actividades diversas, conduzindo a um melhor desempenho da organização.

A integração automática do CCC no conhecimento organizacional apresenta-se como uma potente fonte de conhecimento sistematizada na medida em que permite, por exemplo, recursos para o planeamento, execução e refinamento de campanhas altamente segmentadas e com periodicidade mensal, semanal, diária ou mesmo contínua, bem como verificar a eficácia dessas mesmas campanhas no comportamento dos clientes.

Consideramos, por isso, que, após a aquisição do CCC, deve existir uma fase de actualização e memorização em que os padrões obtidos por DM vão ser assimilados numa **Base de Conhecimento do Comportamento dos Clientes (BCCC)**, permitindo

⁶² Frequentemente, os padrões são fornecidos, a alguns gestores do negócio, sob a forma de relatórios, gráficos ou outros meios semelhantes.

⁶³ Outro perigo que poderá advir para a empresa pelo facto dum gestor a abandonar é que ele é, frequentemente, o único elo de ligação da empresa com certos grupos de clientes os quais vêem a sua relação não necessariamente com a empresa mas mais com o indivíduo em questão; por esse motivo, provavelmente, eles deixarão de ser clientes da empresa.

uma actualização do conhecimento aí existente⁶⁴. Esta base de conhecimento é integrada no conhecimento organizacional a fim de ser utilizada em actividades diversas, conduzindo a um desempenho mais eficiente da organização.

A manutenção de uma base de conhecimento num ambiente dinâmico deve prever um processo de actualização sempre que novas informações sejam obtidas; frequentemente, é problemático assegurar esta actualização, gerando-se, assim, um obstáculo ao sucesso das acções desenvolvidas a partir do conteúdo da referida base.

O facto de não estar implementada, numa empresa, o tipo de prática que se está a propor, pode levar a que os gestores actuem com conhecimento incompleto e desactualizado, não tendo sequer consciência de que houve alterações no comportamento dos seus clientes; mesmo que esse conhecimento esteja actualizado pode não reflectir a sua evolução temporal a qual permite a comparação com factores externos à organização (e.g., políticos, sociais, económicos, meteorológicos e culturais).

4.3 Suporte Tecnológico para Zelar pelo CCC

Na Figura 4.1, está representado um conjunto de ferramentas, oportunamente descritas no capítulo 3, que contribuem para a obtenção e registo do CCC. Com o modelo que é apresentado na Figura 4.2, incluímos dois módulos que consideramos em falta no anterior processo, a fim de que seja possível a actividade de **“Zelar pelo CCC”**, e que são:

- uma base de conhecimento para armazenar o CCC e suportar a sua exploração a que chamamos **Base de Conhecimento do Comportamento dos Clientes (BCCC)**;
- um módulo para a aprendizagem automática e incremental do CCC e verificação da eficácia de acções dirigidas à mudança de comportamento dos mesmos a que chamamos **Gestor de Conhecimento do Comportamento dos Clientes (GCCC)**.

⁶⁴ Algum do conhecimento a memorizar poderá ser obtido a partir de outras fontes que não o resultado do processo de DCBD.

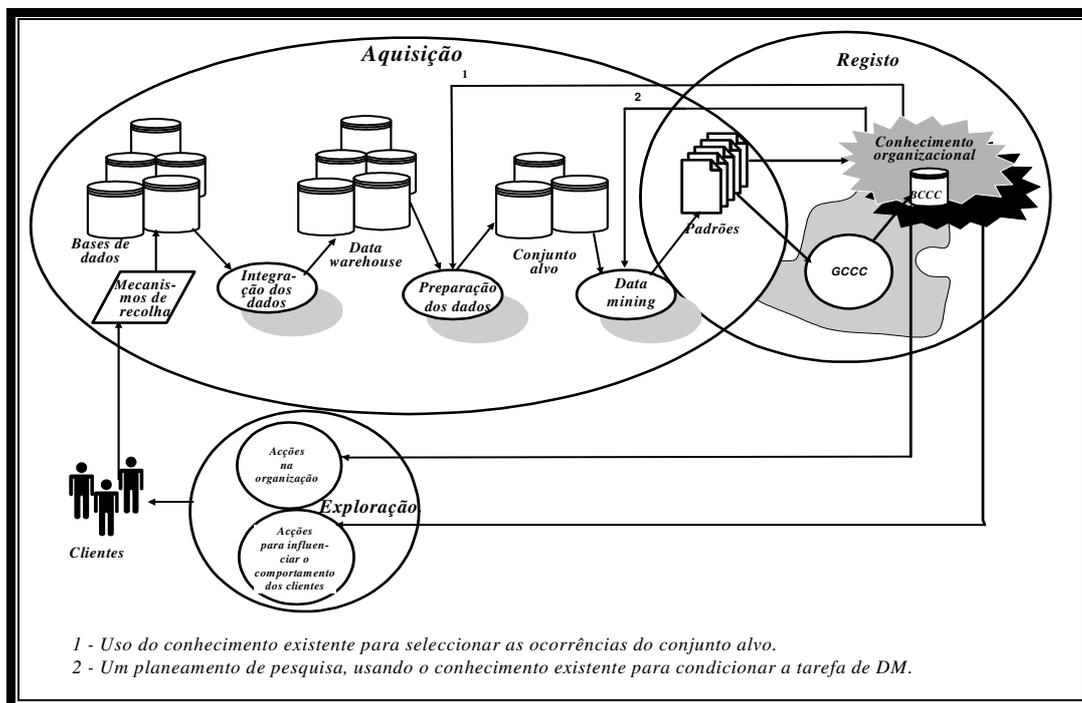


Figura 4.2 Modelo global do processo de aquisição, actualização, memorização e exploração do CCC.

Existindo tecnologia que viabiliza a construção quer da BCCC quer do GCCC, propomo-nos usando essa tecnologia, desenvolver um sistema de conhecimento a que chamamos ZACCAR (Zelar pela Aquisição do Conhecimento dos Clientes, sua Actualização e Registo), suportado nos dois módulos referidos (Figura 4.3).

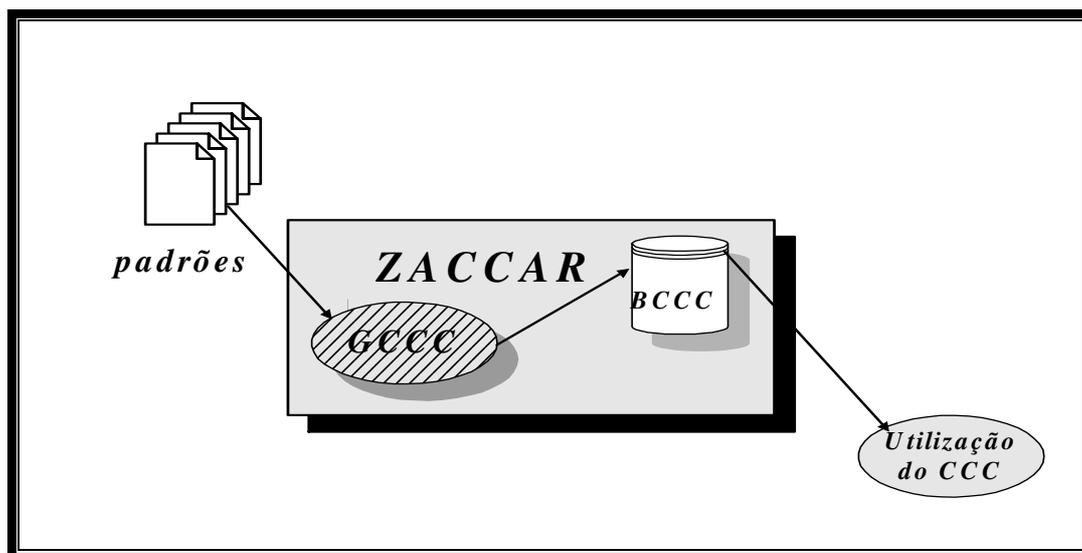


Figura 4.3 – O sistema ZACCAR.

Com a construção deste sistema pretendemos mostrar que, a partir da utilização da tecnologia já existente, é possível chegar-se a soluções que levem a uma melhor condução do negócio.

O sistema ZACCAR, concebido para fazer a ligação entre as bases de dados organizacionais, o conhecimento organizacional e actividades organizacionais directamente ligadas aos clientes, nomeadamente CRM, pode-se posicionar na organização de acordo com o esquema apresentado na Figura 4.4.

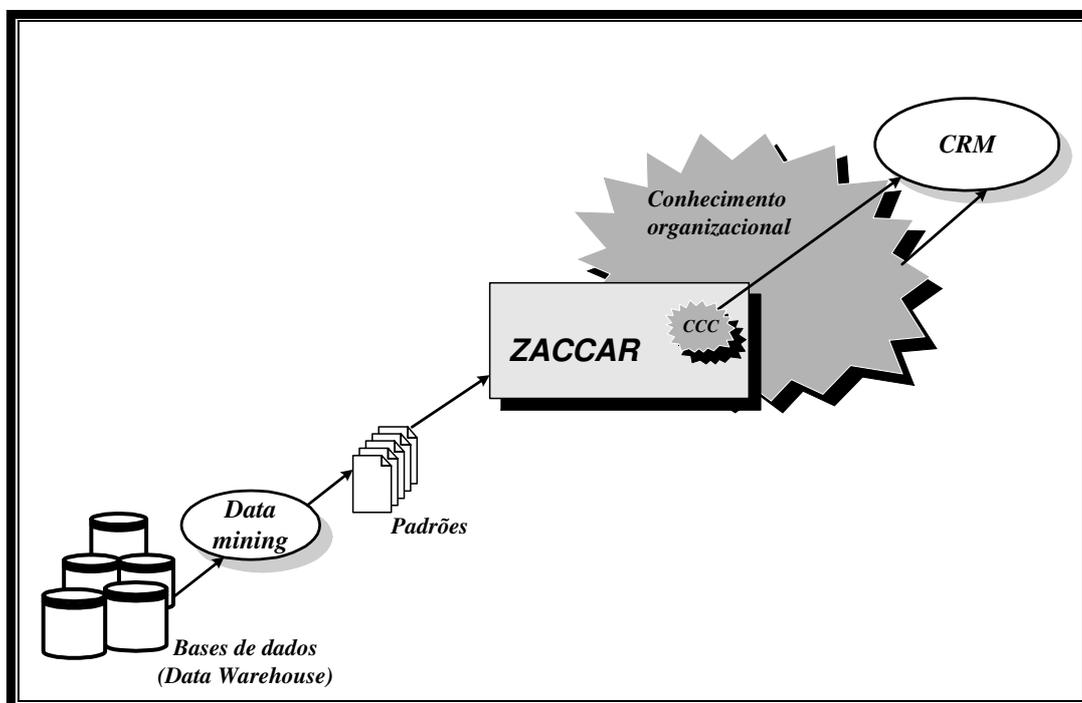


Figura 4.4 - ZACCAR - Uma ponte entre DM e o conhecimento organizacional.

Como se pode observar na Figura 4.4, o sistema ZACCAR vai servir-se do conteúdo das bases de dados (ou do *data warehouse*) da organização, através de ferramentas de DM, para alimentar o CCC o qual faz parte integrante do conhecimento organizacional, contribuindo para as actividades de CRM e, através deste ou directamente, para todas as outras acções que necessitem daquele conhecimento.

4.3.1 A Base de Conhecimento do Comportamento dos Clientes

A BCCC é o repositório de conhecimento onde ficam registados os padrões de comportamento dos clientes, fazendo parte integrante do conhecimento organizacional. O conhecimento aqui registado tem em vista servir de apoio aos gestores para que possam avaliar e caracterizar o mercado para depois actuarem.

Dado o carácter dinâmico das bases de dados operacionais, os padrões obtidos com DM podem variar com muita frequência; há, pois, uma evolução temporal dos padrões que é importante que fique registada. Por esse motivo, a BCCC tem que ser um repositório dinâmico pois, uma vez criado, a sua actualização deve ser constante a partir de novas transacções dos clientes a fim de reflectir as mudanças no conhecimento do domínio.

Atendendo à definição da BCCC apresentada e ao acabado de expor, os requisitos a ter em consideração na criação desta base de conhecimento são:

- armazenar padrões de comportamento;
- associado a cada padrão deverão ser mantidos:
 - ▶ as datas em que os padrões foram detectados;
 - ▶ os valores dos parâmetros níveis de suporte e de confiança em cada uma dessas datas;
 - ▶ comentários do gestor do conhecimento, considerados pertinentes, sobre a evolução do padrão.

A BCCC passará a dispor de um conhecimento actualizado sobre o comportamento dos clientes que servirá, provavelmente, de base ao gestor para tomar melhores decisões em problemas relacionados com o domínio de cada aplicação em concreto e, conseqüentemente, para providenciar bens e serviços mais adaptados aos clientes mais proveitosos para a empresa ou mesmo a potenciais clientes [Freitas & Lavington, 1998].

Uma modificação nos serviços oferecidos conduzirá, muito possivelmente, a uma alteração no comportamento dos clientes a qual se pode traduzir em novas aquisições

mas, também, em novas exigências ou pedidos de esclarecimento. Estas reacções dos clientes ou alterarão directamente as bases de dados transaccionais ou enriquecerão, por outros processos, o conhecimento dos gestores; dum modo ou de outro, essas alterações dever-se-ão reflectir no conteúdo da BCCC.

A frequência do processo de actualização da BCCC depende das áreas do negócio. Assim, em certas organizações como a banca ou os operadores de comunicações móveis em que estão a aparecer, continuamente, novos produtos, há necessidade de actualizações mais frequentes do que acontecerá, provavelmente, noutras organizações em que a inovação dos serviços oferecidos não é tão frequente.

A criação do repositório proposto apresenta numerosos benefícios para a gestão do CCC dos quais se podem destacar:

- regista conhecimento organizacional e disponibiliza-o aos utilizadores específicos na área do negócio, os quais podem usar este conhecimento para desempenhar acções com finalidade na organização ou encarregar-se da sua difusão por outros membros da organização;
- permite uma maior facilidade na comunicação entre os peritos de DCBD e os peritos do negócio, facilitando, por esse motivo, quer o processo de DCBD quer a exploração do CCC⁶⁵;
- conhecimento sobre os clientes que se encontrava disperso e não formalizado⁶⁶ é guardado, após a sua conversão a um formato pré-definido, num único repositório onde pode ser utilizado de maneira útil e eficiente;
- fica registada uma evolução temporal de cada padrão de comportamento;
- o conhecimento é criado e preservado por forma a que não se perca (criam-se condições para que uma memória organizacional fique formalizada e registada,

⁶⁵ Na criação da BCCC, tem que haver uma preocupação na representação do conhecimento, tentando estabelecer um compromisso entre a complexidade necessária para a descrição exacta dos padrões de comportamento e a simplicidade fundamental para a compreensão do utilizador.

independentemente dos gestores em cada momento da vida da empresa ou dos contextos em que as tarefas de DM ou as diversas iniciativas foram executadas);

- são criadas condições para que o conhecimento disponível seja usado (a BCCC pode-se ligar, directamente, às campanhas de *marketing*, sistemas de CRM ou outras aplicações da organização [Moriarty, 1999] e encontra-se estruturada de uma maneira acessível a todos os membros da organização que precisem de utilizar o seu conteúdo, independentemente de terem ou não contribuído para a sua formação);
- permite o desenvolvimento de novo conhecimento, garantindo a actualização frequente da BCCC (constantes aquisições dos clientes, por iniciativa própria ou em função de acções levadas a cabo para modificar o seu comportamento, provocam mudanças frequentes nas bases de dados transaccionais que podem alterar os padrões de comportamento anteriormente registados e, conseqüentemente, o conhecimento obtido a partir deles).

Outras vantagens óbvias da existência deste repositório de conhecimento resultam do facto de que campanhas com periodicidade mensal, semanal, diárias ou até intermitentes podem ser desenvolvidas independentemente do processo de DM que, normalmente, se processará em intervalos de tempo mais alargados ou em situações excepcionais como sejam o período posterior a uma campanha específica de *marketing*.

Uma vantagem emergente da existência desta base de conhecimento é que passa a ser possível não só melhorar os procedimentos habituais mas também viabilizar novas práticas como sejam a possibilidade de aceder à eficácia de campanhas de *marketing* através do confronto do comportamento dos clientes destas acções (reflectidas nos registos das bases de dados da organização) com padrões de comportamento previamente detectados e que se encontram na BCCC.

⁶⁶ Relatórios provenientes da actuação de DM, saber existente na mente dos gestores do negócio e outros suportes documentais.

4.3.2 O Gestor de Conhecimento do Comportamento dos Clientes

O confronto dos padrões obtidos por DM com o conhecimento do perito do negócio e outros padrões já conhecidos bem como a integração do CCC descoberto na BCCC, não pode, devido à sua dimensão, ser feito manualmente. Surgiu, pois, a necessidade de construir um módulo que, possuindo características específicas, desempenhasse aquelas tarefas; chamámo-lo, a este módulo, **Gestor de Conhecimento de Comportamento dos Clientes (GCCC)**.

O objectivo inerente ao seu desenvolvimento, é, sempre que é executado um processo de DM sobre um conjunto alvo:

- registar os padrões obtidos num ficheiro após a sua conversão para um formalismo previamente definido;
- confrontá-los com o conteúdo da BCCC;
- apresentá-los ao gestor do conhecimento organizacional a fim de que os complete;
- integrá-los, já na forma pretendida, na BCCC, através dum processo de assimilação.

Assim, os requisitos a ter em consideração na construção do GCCC são:

- formalizar os padrões de comportamento dos clientes sem que os gestores do conhecimento e os outros utilizadores tenham que lidar com detalhes sintácticos de formalismos de representação do conhecimento;
- fazer uma pesquisa automática de padrões de comportamento;
- consolidar o novo conhecimento, resolvendo potenciais conflitos com o conhecimento já existente;
- actualizar e completar o conhecimento capturado;

- fazer o *interface* com os gestores do conhecimento organizacional⁶⁷, permitindo-lhes a introdução de comentários considerados relevantes, nomeadamente sobre o interesse de determinadas regras [Fayyad *et al.*, 1996a], [Silberschatz *et al.*, 1996], [Silberschatz & Tuzhilin, 1996] ou justificações para alterações observadas nos parâmetros nível de suporte e nível de confiança de uma regra já existente ou, ainda, a remoção de padrões irrelevantes no contexto em causa⁶⁸;
- assimilar o conhecimento actualizado na BCCC.

O interesse do confronto entre os padrões acabados de obter com DM e o conhecimento já existente sobre os clientes e que se encontra previamente formalizado, é óbvio pois que permitirá:

- o registo de novos padrões;
- a actualização de padrões já existentes com os novos parâmetros entretanto obtidos (níveis de suporte e de confiança, intervalos de aplicação para o padrão, novos comentários do gestor do conhecimento);
- a constatação de que um dado padrão se mantém sem alterações;
- a eliminação de padrões obtidos por DM cujo grau de confiança se encontra abaixo dum limite previamente definido o que faz que tenham perdido o interesse como padrões de comportamento.

Sempre que é levado a cabo um confronto, os novos padrões podem ser desconhecidos ou já fazerem parte da BCCC e, neste caso, duas situações se podem verificar:

- os padrões são exactamente iguais;
- o nível de confiança e/ou o nível de suporte da regra representativa do padrão em análise, têm valores diferentes.

⁶⁷ Estes gestores serão, provavelmente, a maior condicionante na implementação da ferramenta pois a sua formação é na área do negócio e tudo o que não lhes for apresentado segundo a sua perspectiva será considerado irrelevante e, portanto, abandonado.

Quando há alteração de um destes parâmetros, a variação detectada pode ser muito significativa; por esse motivo, o gestor do conhecimento organizacional deve acrescentar um comentário que justifique a alteração encontrada. Deste modo, torna-se possível visualizar (ou utilizar para fins convenientes), a qualquer momento, a evolução de cada padrão ao longo do tempo.

Consideramos que a intervenção, neste processo, de um perito do negócio – o gestor do conhecimento - é essencial como a maneira de complementar os padrões descobertos com outra informação já disponível, incluindo comentários na sua justificação e aplicabilidade bem como transmitindo um saber acumulado resultante do seu próprio conhecimento do negócio. É também a ele que caberá definir os limites abaixo dos quais um padrão deixa de ter interesse.

Muitos dos padrões obtidos poderão ser triviais⁶⁹ ou perfeitamente acidentais⁷⁰; é o gestor com o seu conhecimento adquirido em experiências anteriores ou baseado em factores externos e históricos que os eliminará da base de conhecimento.

Além de poder documentar a regra visualizada, o *gestor do conhecimento* deverá, ainda, ter a possibilidade de indicar a realização de tarefas complementares, nomeadamente, repetição do processo de DM, análise directa dos *data warehouses*, a execução de comandos SQL (*Structured Query Language*) sobre os *data warehouses* ou outras tarefas afins.

⁶⁸ Os padrões podem tornar-se irrelevantes, por exemplo quando os parâmetros se mantêm, num certo intervalo de tempo, abaixo de valores previamente definidos.

⁶⁹ São padrões que traduzem comportamento tão evidente que não acrescentam nenhum conhecimento ao já existente.

Exemplo: O aumento da venda de bolo rei na época do Natal.

⁷⁰ Referem-se a uma situação anormal cujas razões poderão merecer uma análise cuidada e que, provavelmente, não se voltará a verificar.

Exemplo: No Verão de 1999, a maior parte das viagens tûristicas para os Açores foram canceladas em virtude dum violento tremor de terra que se fez sentir, no início de Julho, na ilha do Faial.

Deste modo, as razões para a variação dos parâmetros de um dado padrão, poderiam ser detectadas através de uma alteração da estratégia de DM e repetição deste processo ou fazendo *roll-up* ou *drill-down*⁷¹ directamente sobre o conteúdo do *data warehouse* [Berry & Linoff, 1997].

Se se atender a todo o processo de confronto que se verifica na fase da memorização e actualização do conhecimento, facilmente se conclui que as regras armazenadas na BCCC traduzem um profundo conhecimento do domínio o qual tem todas as condições para permitir a definição dum critério fundamental para definir o conjunto alvo e reduzir, por isso, o espaço de pesquisa para o algoritmo de DM com todos os benefícios que daí advêm [Fayyad *et al.*, 1996b] e para ser usado no planeamento da tarefa de DM (ver ligações 1 e 2 na Figura 3.8).

Gera-se, assim, a possibilidade de uma aprendizagem incremental onde o conhecimento previamente gerado e já enriquecido com o conhecimento do gestor do negócio é usado no processo de reconstrução do padrão [Hong & Mao, 1991].

Pelo exposto, pode-se concluir que o GCCC é uma ferramenta de aprendizagem semi-automática sobre o comportamento dos clientes e verificação da eficácia de acções dirigidas à mudança desse mesmo comportamento. É, dados os requisitos a ter em consideração na sua construção, um sistema de aprendizagem incremental e, como tal, tem a capacidade de alterar a BCCC num ambiente dinâmico.

⁷¹ **Roll-up e drill-down** são processos para apresentar a informação em níveis, respectivamente superiores e inferiores dum caminho de integração, entendendo-se este como um nível de agregação em que se pode visualizar a informação [Parsaye & Chignell, 1993], [Lucas, 1997].

4.4 A Viabilidade da Proposta

O demonstrar a viabilidade da proposta apresentada passa por construir protótipos dos módulos propostos e usá-los num determinado contexto.

Num ambiente laboratorial, é possível mostrar a exequibilidade do sistema desenvolvido e detectar situações que carecem de ser melhoradas.

Experimentar o ZACCAR também em ambiente de campo, seria muito útil pois permitir-nos-ia uma análise da sua eficácia sob outros aspectos; mas, dado o horizonte temporal em que nos propúnhamos desenvolver este trabalho e outras condicionantes, nomeadamente de natureza tecnológica, tal não foi possível.

Assim, o que fazemos neste trabalho é:

- identificar uma situação que pode ser problemática;
- apresentar uma proposta para melhorar aquela situação através da introdução dum novo conceito " Gerir o CCC" e estabelecimento de uma forma de estruturar o CCC;
- desenvolver um sistema de conhecimento para a gestão registo do CCC;
- mostrar que o conceito "Gerir o CCC" é viável com recurso a tecnologia existente e construindo um protótipo;
- testar, laboratorialmente, a eficácia do protótipo desenvolvido.

4.5 Sumário

Na sequência da análise da importância, para a empresa, do comportamento dos clientes e o processo da sua obtenção e registo, foi proposta uma nova actividade organizacional: a Gestão do Conhecimento do Comportamento dos Clientes.

Entende-se a gestão do CCC como a actividade de zelar por aquele conhecimento a fim de que, depois da sua criação, seja confrontado com outro conhecimento já

existente, actualizando-o e registando-o, em seguida, numa base de conhecimento a partir de onde pode ser utilizado em actividades diversas na organização.

Todo o processo de zelar por este tipo de conhecimento apresenta, ainda, a vantagem, pelo facto de incorporar uma dimensão temporal, de permitir, em cada momento, a visualização da evolução dum padrão de comportamento ao longo do tempo.

Para tornar realizável a actividade organizacional apresentada, achamos que é necessário o desenvolvimento dum sistema informático - o sistema ZACCAR - composto por:

- uma Base de Conhecimento do Comportamento dos Clientes
- um Gestor de Conhecimento do Comportamento dos Clientes

de que apresentámos os requisitos necessários à sua construção.

Neste capítulo propusemos o sistema ZACCAR e fizemos a sua descrição; no capítulo seguinte, apresentaremos a sua arquitectura, funcionamento e especificação formal.

5.O Sistema ZACCAR

*“Os fins justificam os meios”
(Nicolau Maquiavel, "O Príncipe"⁷²)*

Na sequência da importância que o CCC tem nas organizações propusemos, no capítulo anterior, uma nova actividade organizacional chamada "**Zelar pelo CCC**" e um sistema que consideramos necessário para a sua viabilidade, apresentando uma descrição deste sistema.

A criação de um protótipo do ZACCAR é realizada no sentido de demonstrar a sua exequibilidade e eficiência. A formulação dos requisitos técnicos deste protótipo é apresentada no presente capítulo bem como as opções tecnológicas tomadas para a sua construção.

5.1 Opções tecnológicas

Na escolha das opções tecnológicas (Tabela 5.1), tivemos presente que se pretende que o âmbito de aplicação do ZACCAR deva ser tal que:

- se aplique a um conjunto suficientemente abrangente de domínios de problemas;
- seja mantida a independência com as ferramentas de DM utilizada.

⁷² Nicolau Maquiavel, filósofo, escritor e político italiano (Florença 1469 - 1527)

Tabela 5.1 - Tecnologias utilizadas no desenvolvimento do protótipo.

<i>Tipo de tecnologia</i>	<i>Opção</i>
Sistema operativo	<i>Windows NT</i>
Processador	<i>Intel Pentium</i>
Representação do conhecimento	Regras
Linguagem de programação	<i>Prolog</i>
Ferramenta de <i>data mining</i>	<i>Clementine</i>
Gerador de análise lexical	<i>lex</i> ⁷³
Linguagem hospedeira do compilador	C

5.1.1 Representação do Conhecimento através de Regras de Produção

Um sistema de conhecimento requer um formalismo de representação de conhecimento⁷⁴. O formalismo seleccionado, no contexto deste trabalho, corresponde às regras de produção⁷⁵ [Frost, 1987], [Bratko, 1990] na forma:

se precondição A então conclusão B
se situação A então acção B

em que:

- A é um descritor (ver (3.3)) ou um conjunto de descritores como, por exemplo:
se [A₁ e A₂] ou [A₃ e não A₄]
- B é um descritor.

Esta opção teve em consideração que:

- o modelo de representação de conhecimento baseado em regras de produção está associado à maioria das ferramentas de DM;

⁷³ O *lex* foi desenvolvido nos *Bell Laboratories*, em 1970; desde a versão 7 do *Unix*, o *lex* faz parte dos seus utilitários, sendo a linguagem hospedeira, nas primeiras versões, o C o que facilmente se compreende, dadas as ligações entre esta linguagem e o *Unix*. Actualmente, existem versões do *lex* para ambiente DOS ou *Windows* e o sistema hospedeiro já não é, necessariamente, o C [Aho *et al.*, 1985], [Mason & Brown, 1991].

⁷⁴ Um *formalismo de representação de conhecimento* tem como fim exprimir raciocínio com conhecimento sobre o domínio da aplicação através de uma linguagem bem definida [Brachman & Levesque, 1985], [Brodie & Mylopoulos, 1986], [Frost, 1987], [Stefix, 1995].

⁷⁵ As regras de produção podem admitir informação adicional sob formas variadas [Stefix, 1995].

- a representação do conhecimento deve ser uma representação que ao mesmo tempo que seja eficiente em termos computacionais também seja muito próxima da linguagem natural o que simplifica os requisitos a considerar se houver necessidade de construção dum *interface*;
- uma regra é uma boa escolha pelo facto de que a generalidade das situações que envolve planeamento, diagnóstico, vendas, comportamentos sociais ou outras pode ser expressa, com facilidade, na forma de regras [Dhar & Stein, 1997].

A selecção de regras de produção justifica-se, ainda, pelas suas propriedades de [Bratko, 1990], [Dhar & Stein, 1997]:

- modularidade - cada regra define uma porção de conhecimento;
- incrementabilidade - novas regras podem ser adicionadas à base de conhecimento, independentemente de outras já existentes;
- alterabilidade - as velhas regras podem ser alteradas sem prejuízo das novas;
- transparência - há transparência do sistema de suporte;
- independência
 - ▶ entre regras - as regras mantêm-se independentes umas das outras;
 - ▶ relativamente à utilização - uma regra é guardada sem necessitar de ter em atenção como será usada;
- explicação - torna-se fácil responder a questões como:
 - ▶ como chegou a esta conclusão? (Questões COMO);
 - ▶ porque está interessado nesta informação? (Questões PORQUÊ);
- separação lógica - há uma separação entre o quê e o como da regra;
- interpretação - é fácil, para o utilizador, compreender e validar o conhecimento expresso na regra.

Claro que nalgumas situações, a utilização de regras para representação do conhecimento pode não ser aconselhável, nomeadamente, se as regras se tornam demasiado extensas ou se o número de regras necessário à representação do conhecimento é excessivamente grande.

Houve necessidade de que as regras de produção utilizadas no sistema de conhecimento desenvolvido fossem estendidas para que, além do formalismo anterior, registassem, ainda, informação complementar.

A sintaxe destas regras em notação EBNF⁷⁶ (*Extended Backus Naur Form*) (os símbolos são explicados na Tabela 5.2) [Howe, 1995], [Garshol, 2000] corresponde a:

```

<regra> ::= regra se <descriptorA> então <descriptorB>
           <extensão>

<descriptorA> ::= <terno> { <op_booleano> <terno> } *
<descriptorB> ::= <terno>
<op_booleano> ::= e | ou
<terno> ::= <atributo> <operador> <valor>
<atributo> ::= <letra> | <dígito> { <letra> | <dígito> | _ } *
<operador> ::= < | > | <= | >= | = | ≠ | ∈ | ∉
<valor> ::= <inteiro> | <real> | <string>
<extensão> ::= parâmetros [ <níveis> ]
              | evolução [ <momento> { , <momento> } * ]
<níveis> ::= n1 = <real> , n2 = <real>
<momento> ::= ( <data> , <comentário> , <parâmetros> )
<data> ::= data ( <dígito> <dígito> , <dígito> <dígito> , <dígito> <dígito> <dígito> <dígito> )
<comentário> ::= comentário (<string>)
<parâmetros> ::= parâmetros (<níveis>)
<dígito> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
<inteiro> ::= [-|+] <inteiro sem sinal>
<inteiro sem sinal> ::= <dígito> +
<real> ::= <inteiro> . <inteiro sem sinal> [ E <inteiro> ] | <inteiro> E <inteiro>
<letra> ::=
A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q
r|s|t|u|v|w|x|y|z
string - qualquer conjunto de caracteres iniciados com ' e terminados em ' e sem ' no interior.

```

Tabela 5.2 - Simbologia EBNF utilizada na sintaxe das regras.

Símbolo	Denota
::=	é definido como
<...>	regra sintáctica
	ou
{ op ₁ op ₂ ... }	ocorre uma ou mais das opções
[...]	ocorre no máximo uma vez
+	ocorre pelo menos uma vez
*	ocorre zero vezes ou mais

Entre elas, consideramos dois tipos:

1. **Regra Canónica** é uma regra que representa o conhecimento oriundo das ferramentas de DM que sofreu um processo de transformação a fim de ficar registado com o formalismo seguinte:

```
regra se <descriptorA>  
então  
<descriptorB>  
parâmetros  
[ $n_1, n_2$ ]
```

em que n_1 e n_2 determinam o grau de incerteza da regra e têm o significado seguinte:

- n_1 é a percentagem dos elementos do conjunto observado que satisfaz a *descriptorA* (nível de suporte);
- n_2 é a percentagem dos elementos que, satisfazendo a *descriptorA* também satisfazem a *descriptorB* (nível de confiança) [Simoudis, 1996].

A regra canónica, tal como a formulamos, apresenta o valor dos parâmetros obtidos numa única observação.

2. **Regra Canónica Estendida** é uma regra canónica à qual se acrescenta a sua evolução ao longo da coordenada tempo e toma-se amostras em n momentos correspondentes a ciclos de vida da organização (e.g., lançamento de um novo produto, período após uma campanha de *marketing*). Formalmente:

```
regra se <descriptorA>  
então  
<descriptorB>  
evolução  
[( $data_n$ , comentário $_n$ , parâmetros $_n$ )  
...  
( $data_2$ , comentário $_2$ , parâmetros $_2$ )  
( $data_1$ , comentário $_1$ , parâmetros $_1$ )]
```

⁷⁶ A notação EBNF foi normalizada através do documento ISO/IEC 14977:1996 (E).

em que:

- $data_i$, $i=n, \dots, 1$ representam as datas de actualização da regra;
- $comentario_i$ é o comentário do gestor do conhecimento na $data_i$, $i=n, \dots, 1$;
- $parâmetros_i$ determina o grau de incerteza da regra na $data_i$, $i=n, \dots, 1$, sendo composto por:
 - ▶ $suporte_i$, o nível de suporte da regra;
 - ▶ $confiança_i$, o nível da confiança da regra.

As regras canónicas estendidas, geradas pela actuação do ZACCAR sobre as regras canónicas, correspondem ao formalismo utilizado para memorizar o CCC na BCCC; podemos observar que cada uma destas regras contém:

- o padrão descoberto, acompanhado dos respectivos parâmetros indicativos dos níveis de suporte e confiança;
- as datas de integração do conhecimento com o fim de ser possível, em qualquer momento, atendendo a uma dimensão temporal que fica, então, registada, traçar a evolução do comportamento dos clientes;
- justificações para o aparecimento de novos elementos de conhecimento ou modificações significativas nos parâmetros associados ao conhecimento existente;
- comentários sobre o interesse dos padrões de acordo com o parecer do perito do negócio⁷⁷.

Como se pode observar na especificação destas regras, a regra sintáctica <descriptorA> pode representar diferentes situações como sejam:

- uma relação de igualdade como, por exemplo:

*regra se actividade=investigação
então*

⁷⁷ Este campo de comentários permite que o conhecimento do perito do negócio resultante dum saber acumulado e que não se pode exprimir num formato "SE-ENTÃO" seja também registado e utilizado em acções posteriores no negócio. O seu interesse, dado o conhecimento de que é detentor aquele perito, contribui tanto para o conhecimento a preservar como a parte restante da regra, formalmente obtida. Há a necessidade de estruturar a forma como se desenvolve esse conhecimento para que cada um possa entendê-lo sem ambiguidade.

página de acesso=www.amazon.com

- uma conjunção de várias relações de igualdade como, por exemplo:

regra se actividade=investigação e página de acesso=www.altavista.com e página de acesso=www.yahoo.com
então página de acesso=www.amazon.com

- um relação de inclusão como, por exemplo:

regra se idade \in [16, 25]
então página de acesso=www.napster.com

Com esta opção para a representação do conhecimento pretendemos que o protótipo desenvolvido seja independente das situações representadas nas regras por $\langle \text{descriptorA} \rangle$ e $\langle \text{descriptorB} \rangle$; assim, em cada caso particular, estas regras sintácticas são automaticamente concretizadas o que irá permitir uma grande abrangência nos domínios de problemas a tratar.

Cada uma das regras assim constituídas é encarada como um objecto de conhecimento a utilizar posteriormente em actividades diversas na organização e que é mantido e gerido como quaisquer outros dados da organização, fazendo parte integrante dos recursos de dados organizacionais e, portanto, do conhecimento organizacional da empresa.

5.1.2 A linguagem de programação em lógica *Prolog*

Tendo em atenção o formalismo escolhido para a representação do conhecimento na BCCC (representação puramente declarativa) e uma ampla abrangência de domínios de problemas em que possa vir a ser aplicado o sistema desenvolvido, tornou-se evidente que a escolha da linguagem tinha que tender para uma linguagem de programação em lógica⁷⁸.

⁷⁸ Um *programa lógico* é a especificação formal dum problema; é composto por um conjunto de axiomas ou regras, definindo relações entre objectos. Os programas possuem uma estrutura modular uma vez que os predicados constituem um bloco de conhecimento independente e autocontido [Santos, 1999].

A opção adoptada foi o *Prolog*⁷⁹ [Bratko, 1990], [Dougherty, 1994], [Cardoso, 1995], [Santos, 1999], pois que:

- é uma linguagem orientada para o manuseamento de regras e factos, permitindo um mais fácil desenvolvimento de mecanismos de inferência;
- é possível registar o conhecimento com um aspecto muito próximo da linguagem natural, usando termos específicos do domínio da aplicação;
- permite uma documentação automática dos programas através da possibilidade de registo das regras, usando frases e termos da linguagem corrente;
- basta declarar o modelo do problema para que o motor de inferência pesquise uma solução uma vez que esta é uma linguagem declarativa baseada na lógica de predicados.

Além dos argumentos apresentados, o *Prolog* é uma boa linguagem para o desenvolvimento de protótipos, permitindo uma criação rápida de programas exploratórios que podem ser experimentados e testados, possibilitando, assim, o desenvolvimento incremental do protótipo em que o código se vai enriquecendo à medida que vai sendo experimentado.

Enquanto nas linguagens de programação de alto nível (e.g., *C*, *Pascal*, *Algol*), se programa o "como se faz", sendo pois necessário exprimir, rigorosamente, e dum modo objectivo, todos os passos da resolução do problema, em programação em lógica programa-se "o que fazer", cabendo, à linguagem, a preocupação do "como"⁸⁰.

Assim, cada problema que se pretende programar apenas tem que ser representado convenientemente, através da especificação dum conjunto adequado de regras de inferência.

Várias opções são possíveis na escolha de uma linguagem de programação lógica mas o *Prolog* é a que maior sucesso teve devido às suas técnicas de compilação

⁷⁹ **PROLOG** = *Programming in Logic*.

⁸⁰ A arte de programação em lógica consiste na construção de programas concisos e elegantes que cumprem o objectivo para que são desenvolvidos [Sterling & Shapiro, 1994].

sofisticadas; é, provavelmente por esse motivo, destas linguagens, a única que vingou comercialmente. [Clocksin & Mellish, 1987], [Coelho & Cotta, 1988], [Bratko, 1990], [Scott & Nicolson, 1991], [Dougherty, 1994], [Marciniak, 1994], [Shoam, 1994], [Lugger & Stubblefield, 1998].

O *Prolog* caracteriza-se por ser uma linguagem baseada em cláusulas de Horn⁸¹ utilizando técnicas que combinam a resolução directa com o encaminhamento para trás (*backward chaining*) como procedimento de prova (dedução) [Santos, 1999].

Aproveitando capacidades específicas do *Prolog*, houve a preocupação, na construção do protótipo, de fazer algumas opções que o tornassem mais eficiente e permitissem uma melhor legibilidade do código desenvolvido (Anexo 3).

Assim, além da utilização dos operadores reconhecidos por esta linguagem, como sejam o “=”, o “<”, ou o “>”, acrescentámos os operadores "*regra*", "*se*", "*entao*", "*em*", "*parametros*" e "*evolucao*" com a sintaxe⁸² que se apresenta:

```
:-op(850,fx,regra).
:-op(750,xfx,entao).
:-op(720,fx,se).
:-op(710,xfy,e).
:-op(705,xfx,em).
:-op(860,yfx,[evolucao,parametros]).
```

A utilização destes construtores sintácticos permite a representação das regras numa forma perfeitamente compreensível por um qualquer agente organizacional como, facilmente, se pode verificar nos exemplos a seguir apresentados:

Exemplo 1: *regra se idade<19.5 e area_residencia = `C` e agregado_familiar > 3.5*
(regra *entao*
canónica) *opcao = `S`*
parametros
[n1=1.49,n2= 67]

⁸¹ *Cláusulas de Horn* são asserções lógicas na forma normal conjuntiva com um literal positivo no máximo [Shinghal, 1992].

⁸² Um operador em *Prolog* é definido por `:-op(precedência, tipo, nome)`, podendo o tipo representar um operador *infix* (xfx, xfy, yfx), um operador *prefix* (fx, fy) ou um operador *postfix* (xf, yf) [Bratko, 1990].

Exemplo 2: *regra se idade < 19.5 e area_residencia = 'C' e agregado_familiar > 3.5*
(regra *entao*
canónica *opcao = 'S'*
estendida) *evolucao*
[(data (30,01,2001), comentario('Depois da campanha de marketing 3 – Varadero+Havana, pago a prestações `), parâmetros (n1=1.49, n2=67)),
(data (30,12,2000), comentario('Depois da campanha de marketing 2 – Varadero,
pago a prestações `), parâmetros (n1=1.65, n2=25)),
(data (30,11,2000), comentario('Depois da campanha de marketing 1 – Varadero+Havana`), parâmetros (n1=4.202, n2=73)),
(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)`),
parâmetros(n1=4.51,n2=35))].

Uma outra característica poderosa do *Prolog* relaciona-se com o processo de unificação de termos.

5.2 A Arquitectura Funcional do Sistema

O sistema proposto é, como já se viu no capítulo 4, composto por:

- um **G**estor de **C**onhecimento do **C**omportamento dos **C**lientes (GCCC);
- uma **B**ase de **C**onhecimento do **C**omportamento dos **C**lientes (BCCC);
- um módulo de consulta à BCCC.

Desenvolvemos, por razões óbvias, um módulo de consulta à BCCC, permitindo, ao consumidor do conhecimento (e.g., um membro da organização, um programa informático) o acesso e pesquisa dos padrões de comportamento registados naquela base.

A nossa preocupação foi a de demonstrar a exequibilidade do sistema; assim, algumas características foram tidas em consideração no desenvolvimento deste sistema, nomeadamente:

- possibilidade de estruturação e conceptualização do conhecimento sem necessidade de adquirir conhecimento aprofundado sobre o domínio do problema;
- integração automática no conhecimento organizacional;

- interactividade, permitindo o relacionamento com o utilizador humano ou com outros sistemas informáticos;
- modularidade, permitindo a sua integração noutros sistemas organizacionais;
- simplicidade de utilização.

Na Figura 5.1 é apresentada a arquitectura do sistema desenvolvido, seguindo-se uma descrição de cada um dos seus constituintes.

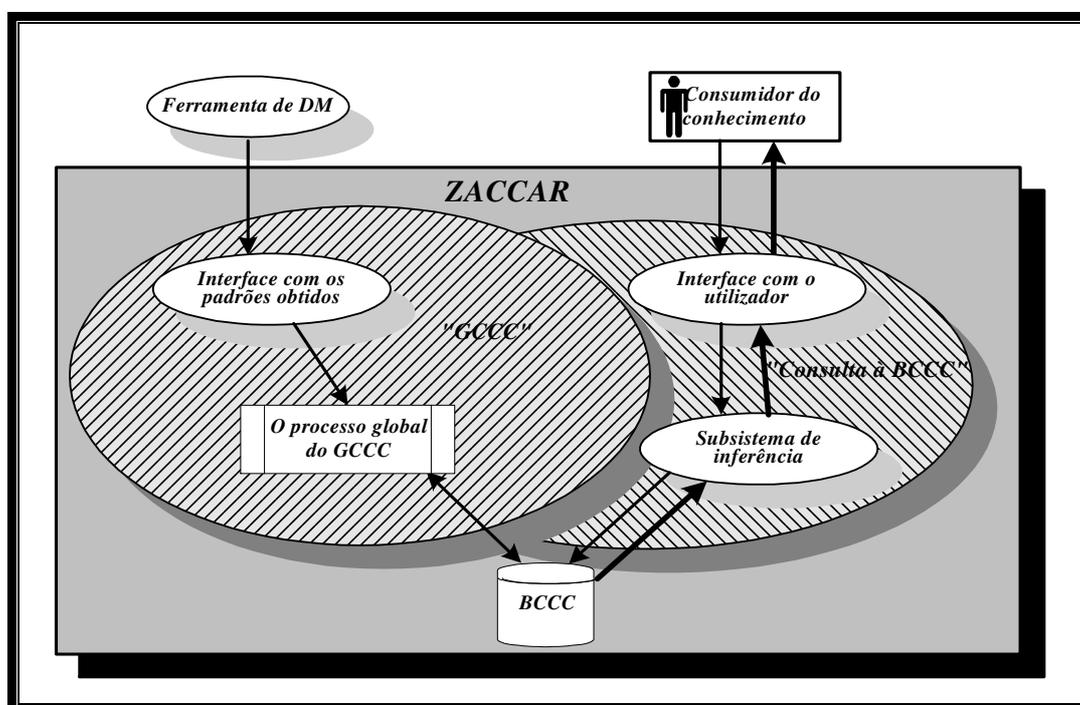


Figura 5.1 – Arquitectura do sistema de conhecimento ZACCAR.

5.2.1 O Gestor de Conhecimento do Comportamento dos Clientes.

O GCCC permite que o conhecimento descoberto pelo processo de DM seja integrado no conhecimento organizacional através do seu registo numa base de conhecimento, após ser confrontado com outro conhecimento aí registado, resolvendo possíveis conflitos e incluindo, também, uma evolução temporal.

Este processo conforme é dissecado nas próximas secções, desenrola-se ao longo de quatro fases (Figura 5.2) que são:

- aquisição do conhecimento;
- confronto do conhecimento;
- participação do gestor do conhecimento;
- assimilação do conhecimento.

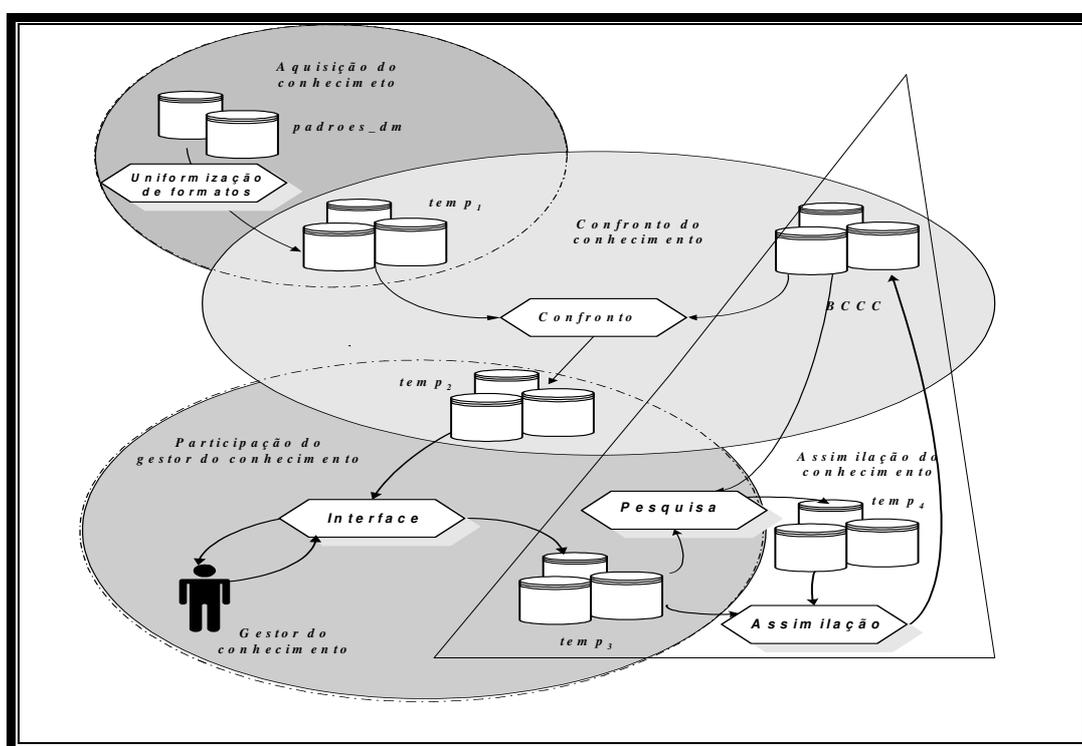


Figura 5.2 - O processo do global do GCCC e respectivas fases

5.2.1.1 Aquisição do conhecimento

Por aplicação de uma ferramenta de DM, actuando sobre os dados registados em bases de dados transaccionais ou no *data warehouse* construído a partir daquelas bases de dados, obtém-se padrões de comportamento que são registados, sob a forma de regras, num ficheiro de texto.

Por forma a manter a independência com as ferramentas de DM utilizadas, a primeira fase de execução do GCCC tem como fim uniformizar os formatos dos padrões

obtidos, transformando-os em regras canónicas que regista no ficheiro *temp₁* (Figura 5.2). Para tal, foi desenvolvido um procedimento chamado “Uniformização de Formatos”, ao qual compete a transformação de uma representação de conhecimento noutra.

Esta transformação dos padrões de comportamento, obtidos com uma ferramenta de DM, representados num qualquer formalismo de representação de conhecimento em regras canónicas garante que as regras que vão ser armazenadas na BCCC estejam sintacticamente coerentes.

Este procedimento formaliza-se, matematicamente, pela função *unif* definida por:

$$\begin{aligned} \text{unif} : F_1 &\rightarrow F_2 \\ r_1 &\mapsto r_2 \end{aligned}$$

com:

- $\text{unif}(r_1) = r_2$ a regra canónica correspondente à regra r_1 obtida com a ferramenta de DM;
- $F_1 = \{\text{regras obtidas com a ferramenta de DM}\}$;
- $F_2 = \{\text{regras canónicas}\}$.

Os conjuntos F_1 e F_2 correspondem, respectivamente, aos ficheiros *padroes_dm* e *temp₁* (Figura 5.2).

Para cumprir o procedimento “Uniformização de formatos”, foi utilizado um compilador gerado pelo *lex*, baseado numa gramática; esta gramática é construída para cada ferramenta (ou algoritmo) de DM utilizada pelo facto de diferentes ferramentas de DM poderem gerar padrões de comportamento com sintaxes diferentes.

O compilador gerado materializa a função *unif*.

A primeira vez que uma determinada ferramenta de DM é utilizada, um procedimento é executado, tendo como objectivo a criação dum programa executável – o compilador - a ser integrado no ZACCAR. Este procedimento (Figura 5.3) consta de:

- 1) criar uma gramática, definida em função da sintaxe associada à ferramenta (ou algoritmo) de DM, que é guardada num ficheiro de texto a que chamámos *tabela.lex*⁸³;
- 2) gerar o compilador, programa na linguagem hospedeira, usando o comando *lex tabela.lex*;
- 3) compilar o programa obtido.

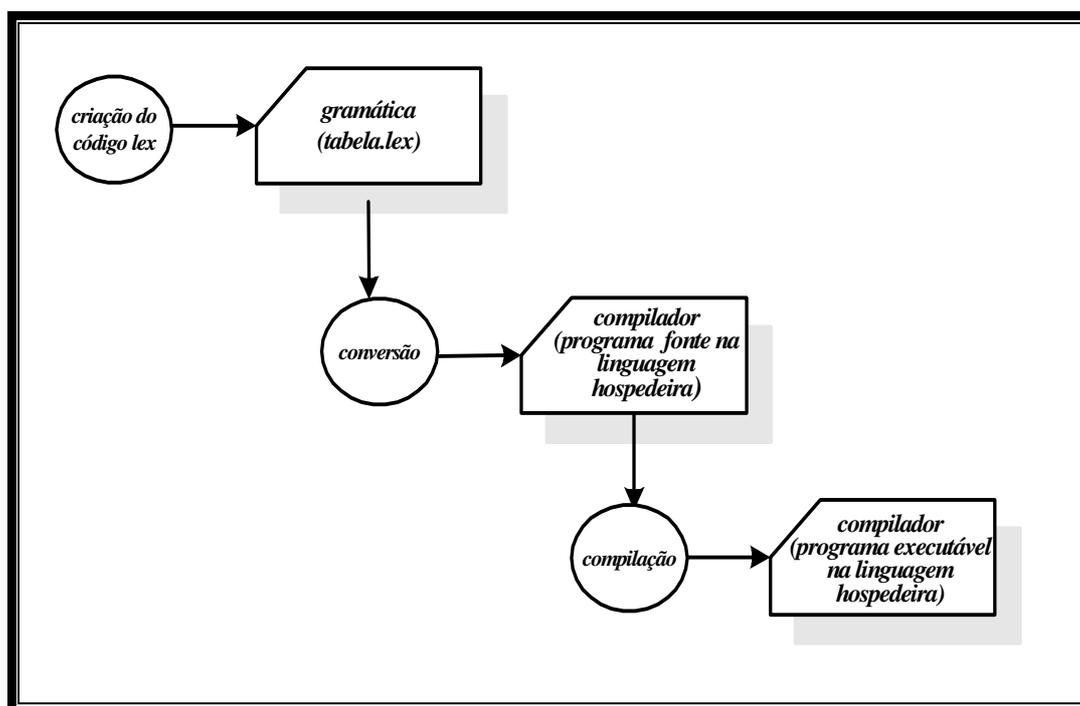


Figura 5.3 - Procedimento para gerar o compilador.

Uma vez criado o compilador, o ZACCAR tem condições para executar o procedimento "Uniformização de formatos" (Figura 5.4), traduzindo o conteúdo do ficheiro *padroes_dm* para regras canónicas que regista no ficheiro *temp1*.

⁸³ A tabela *lex* tem que ser padronizada de acordo com o conteúdo do ficheiro *padroes_dm* (Figura 5.2), atendendo a aspectos sintácticos e semânticos associados à ferramenta de DM que se utilize.

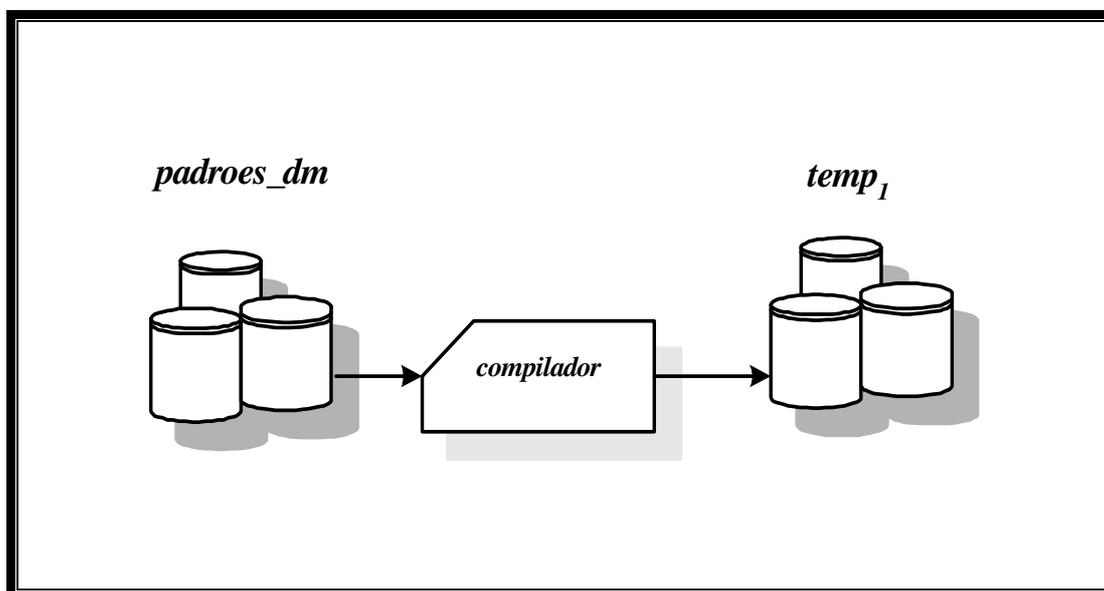


Figura 5.4 - Procedimento "Uniformização de formatos".

5.2.1.2 Confronto do conhecimento

Esta fase é a que desempenha a actividade central em todo o processo. Consiste no confronto do novo conhecimento com o conhecimento já existente e que se encontra registado na BCCC, actualizando este ou integrando os novos padrões.

Para a sua viabilidade, foi desenvolvido o procedimento “Confronto” (Figura 5.5) (Anexo 3.1).

Para que fosse possível o confronto, introduzimos o conceito de *regras semelhantes*, i.e., regras em que os antecedentes são unificáveis assim como os consequentes.

Conceptualmente, há, neste procedimento, três tarefas:

- a pesquisa que corresponde à procura, na BCCC, dum padrão de comportamento com o mesmo consequente e os mesmos antecedentes do novo padrão;
- a acção que corresponde à actualização dum padrão já existente com os novos parâmetros (data da ocorrência e os valores dos níveis de suporte e de confiança) ou à transformação dum novo padrão de modo a poder ser representado por uma regra canónica estendida

- o registo, no ficheiro $temp_2$, da regra correspondente ao padrão depois de actualizado.

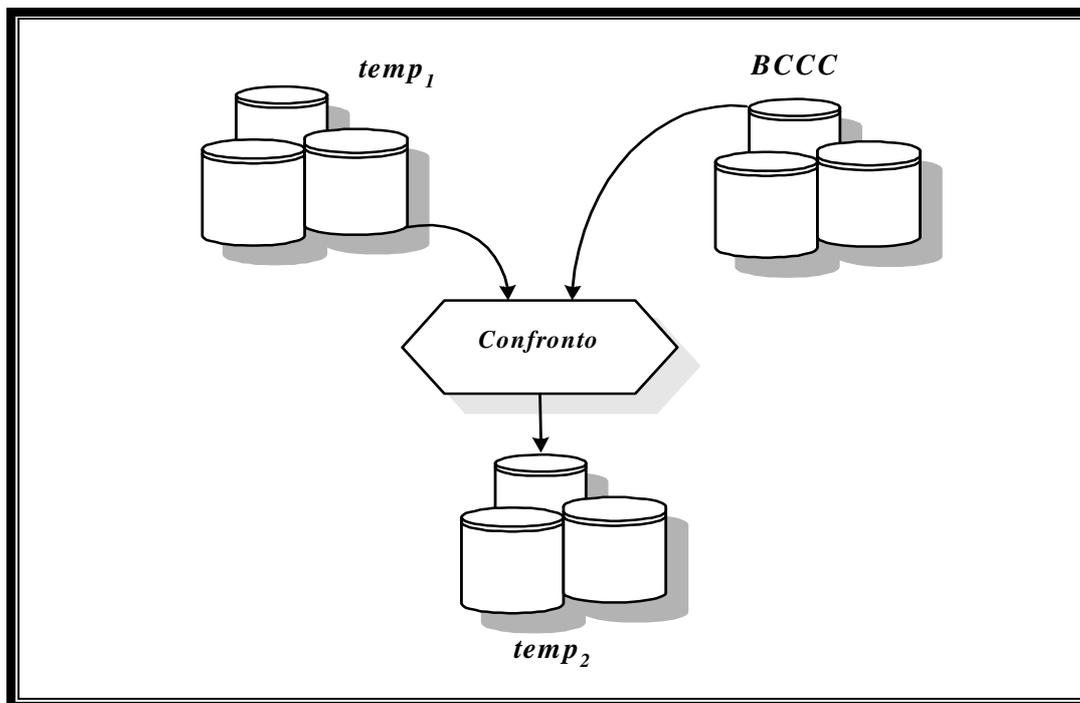


Figura 5.5 – Procedimento “Confronto”.

O algoritmo para a implementação deste procedimento é em seguida descrito:

```

enquanto  $temp_1$  não vazio fazer
  trata próxima regra
  pesquisa                               (*consulta a BCCC*)
  pesquisa regras com conseqüente igual ao da regra a tratar
  se pesquisa com sucesso
  então
    pesquisa das regras seleccionadas a que tem antecedentes iguais aos da regra a tratar
    se pesquisa com sucesso
    então
      as regras são semelhantes
    senão
      a regra em análise é nova
    fimse
  senão
    a regra em análise é nova
  fimse
  acção                                 (*actualização*)
  
```

se regras semelhantes

então

na evolução da regra encontrada é inserida a nova ocorrência (data e valores parâmetros de suporte e confiança)

senão

a nova regra é transformada numa Regra Canónica Estendida.

fimse

*registo (*a regra actualizada é escrita no ficheiro temp₂*)*

apaga a regra tratada

fimenquanto

elimina temp₁

Formalmente, este procedimento é definido com base em duas funções $conf_1$ e $conf_2$ de domínios diferentes mas definidas sobre o mesmo conjunto imagem e definidas, respectivamente, por:

$$- conf_1 : F_2^1 \times F \rightarrow F_3$$

$$(r_2, r) \mapsto r \oplus r_2$$

em que:

- $F = \{\text{regras canónicas estendidas que constituem a base de conhecimento}\};$
- $F_2 = \{\text{regras canónicas}\};$
- $F_3 = \{\text{regras canónicas estendidas actualizadas no momento presente}\};$
- $F_2^1 \subset F_2 = \{\text{regras canónicas a que correspondem regras semelhantes em } F\};$
- \oplus é o operador que actualiza a regra r com os parâmetros de r_2 e a data da actualização.

$$- conf_2 : F_2^2 \rightarrow F_3$$

$$r_2 \mapsto r_3$$

em que:

- $conf_2(r_2) = r_3$ é a regra canónica estendida correspondente à regra canónica r_2 ;
- $F_2^2 = F_2 \setminus F_2^1$;
- \setminus é o operador diferença de dois conjuntos.

Enquanto a função $conf_1$ actualiza regras já existentes com os parâmetros correspondentes a uma regra semelhante que está a ser confrontada no momento, a função $conf_2$ actua só sobre novas regras, transformando-as em regras canónicas

estendidas. As imagens resultantes, nas duas situações, são regras canónicas estendidas e o conjunto imagem é sempre F_3 .

Como se pode facilmente observar, há uma correspondência biunívoca, respectivamente, entre os conjuntos F , F_2 e F_3 e os ficheiros $BCCC$, $temp_1$ e $temp_2$ apresentados na Figura 5.5.

5.2.1.3 Participação do gestor do conhecimento

Esta fase representa a parte não automatizada do processo pois é a única em que é solicitada a interferência humana; todo o restante processo é perfeitamente autónomo e transparente ao utilizador da ferramenta.

Nesta fase, o gestor do conhecimento intervém na actualização das regras ou eliminação daquelas que já são consideradas sem interesse para a empresa; esta participação permite a captura, para o sistema, do conhecimento do perito do domínio.

O gestor do conhecimento, face à análise da evolução da regra que lhe é apresentada e atendendo ao seu conhecimento do negócio, tem a possibilidade de acrescentar justificações para essa evolução ou outros comentários que considere pertinentes, nomeadamente sobre campanhas levadas a cabo antes do momento que está em análise.

Além disso, pelo facto de alguns padrões se poderem tornar obsoletos ou não terem interesse no contexto do negócio, é dada a possibilidade da eliminação das respectivas regras. Assim, esta participação do gestor do conhecimento garante, pelo facto de lhe ser facultada esta possibilidade, que as regras sem interesse ou sem sentido não sejam integradas na BCCC.

Para a viabilidade desta fase foi desenvolvido o procedimento “*Interface*” (Figura 5.6) (Anexo 3.2).

Após este procedimento, as regras actualizadas, caso não tenham sido eliminadas, são guardadas num ficheiro temporário que designamos por $temp_3$.

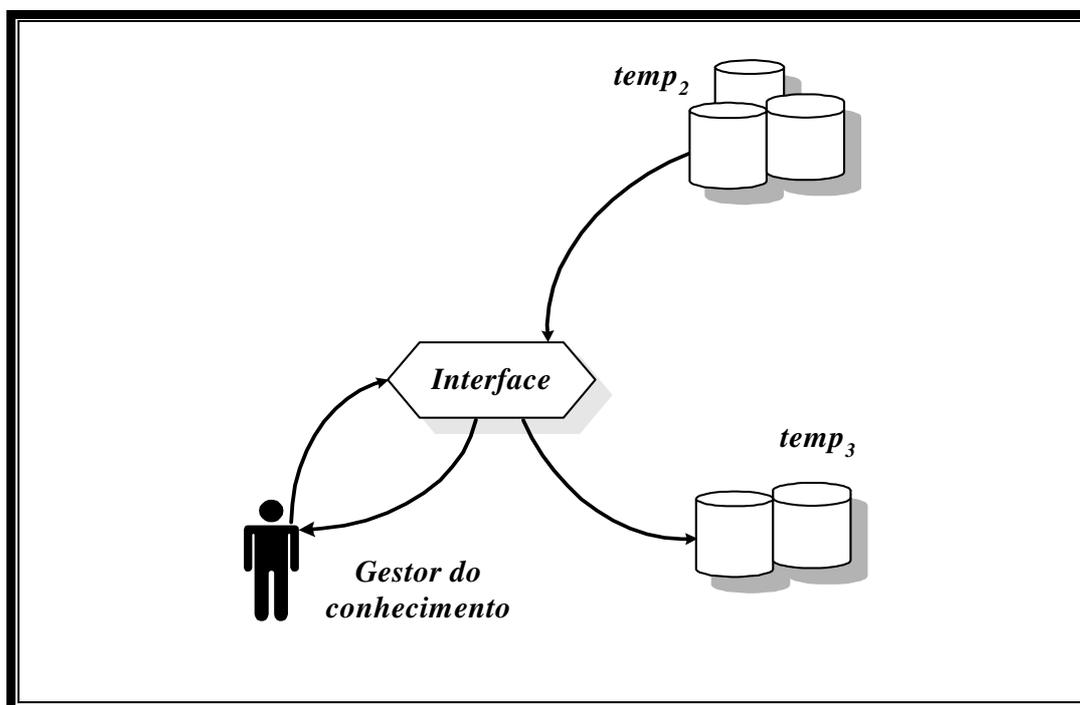


Figura 5.6 - Procedimento “Interface”.

Conceptualmente, consideramos 3 tarefas (Tabela 5.3) para cada padrão apresentado ao gestor do conhecimento:

- opção: a apresentação das possíveis opções (Figura 5.7), permitindo a sua selecção;
- reacção: a acção desencadeada quando uma das opções é seleccionada;
- registo: a possibilidade de registo, em $temp_3$, do padrão depois de actualizado.

Tabela 5.3 - Tarefas do procedimento “Interface”.

<i>opção</i>	<i>acção</i>	<i>registo</i>
1	O padrão é eliminado	Não
2	Digitação de comentário	Sim
3	Inserção de frase	Sim
4	Inserção de frase	Sim
5	Digitação de comentário	Sim

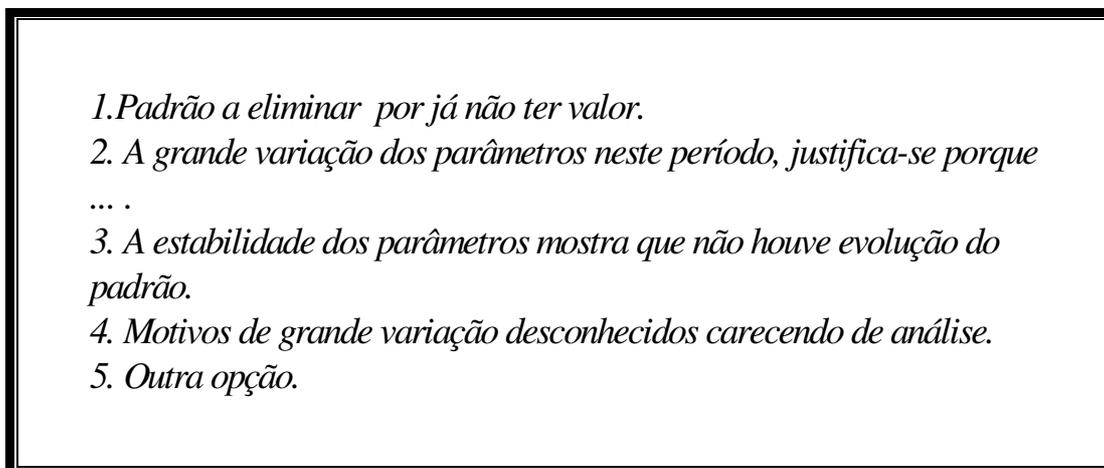


Figura 5.7 - Opções a serem seleccionadas pelo gestor do conhecimento.

O procedimento “*Interface*” é formalizado, matematicamente, pela função *interface* definida por:

$$- \text{interface} : F_3^I \rightarrow F_4$$

$$r_3 \mapsto r_3 \bullet k$$

em que:

- $F_3 = \{\text{regras canónicas estendidas actualizadas no momento presente}\};$
- $F_4 = \{\text{regras canónicas estendidas incluindo comentário do gestor do conhecimento}\};$
- $F_3^I \subset F_3 = \{\text{regras que o gestor do conhecimento não pretende eliminar}\};$
- k é o comentário seleccionado pelo gestor do conhecimento;
- r_3 é uma regra;
- \bullet é o operador que insere na regra r_3 o comentário k .

Como se pode facilmente observar, os conjuntos F_3 e F_4 utilizados na definição da função *interface*, correspondem, respectivamente, aos ficheiros *temp₂* e *temp₃* (Figura 5.6).

5.2.1.4 Assimilação do conhecimento

Ao ser atingida esta fase, existem na BCCC dois conjuntos de regras: as regras que foram seleccionadas para serem actualizadas na sessão presente e cujo registo actualizado se encontra no ficheiro *temp₃* (onde também se encontram as novas regras) e as que o não foram. Há que assimilar todo este conhecimento para a BCCC a fim de ser utilizado em actividades diversas na organização.

A viabilidade desta fase é conseguida por dois procedimentos: “Pesquisa” e “Assimilação” (Figura 5.8) (Anexo 3.3).

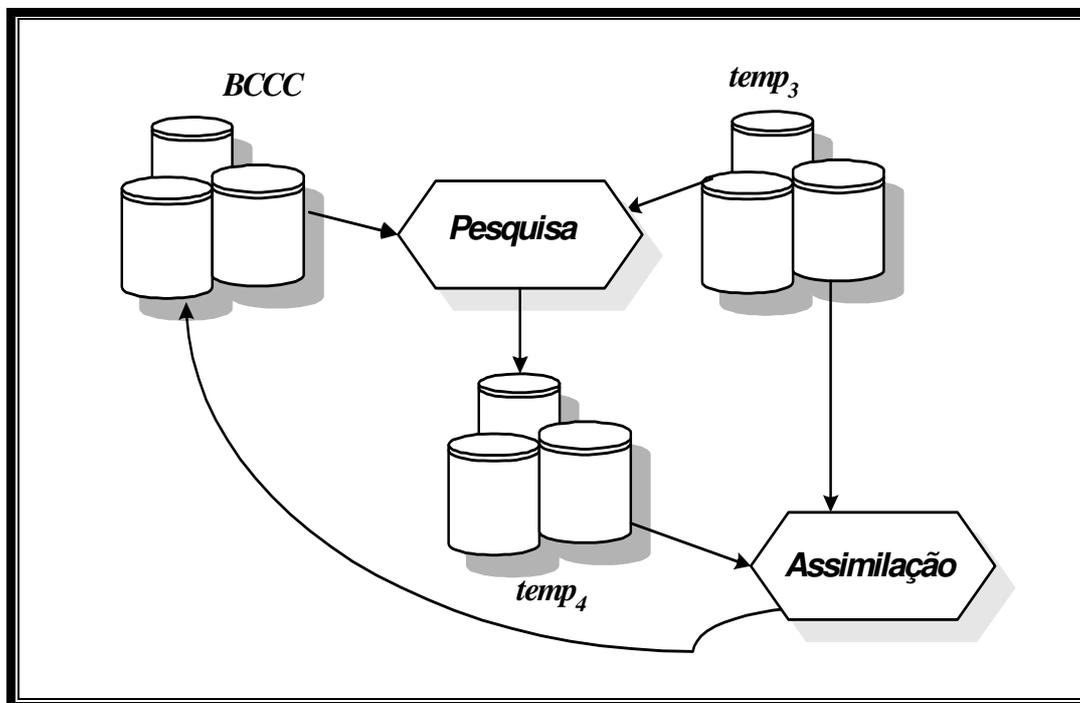


Figura 5.8 - Os procedimentos "Pesquisa" e "Assimilação"

Conceptualmente, o procedimento “Pesquisa”, para cada padrão registado na BCCC (Figura 5.8), cumpre-se com duas tarefas (Tabela 5.4):

- pesquisa dum padrão, em *temp₃*, semelhante ao padrão em análise;
- possibilidade de registo do padrão em *temp₄*.

Tabela 5.4 - Tarefas no procedimento "Pesquisa".

<i>pesquisa</i>		<i>registo</i>
<i>consequente</i>	<i>antecedente</i>	
Sim	Sim	Não
Sim	Não	Sim
Não	--	Sim

Formalmente, este procedimento corresponde a:

$$- \textit{pesquisa} : F^I \rightarrow F_5$$

$$r \mapsto r$$

em que:

- $F = \{\text{regras canónicas estendidas que constituem a base de conhecimento}\};$
- $F_5 = \{\text{regras canónicas estendidas não actualizadas no momento presente}\};$
- $F_4 = \{\text{regras canónicas estendidas actualizadas no momento, incluindo já comentário do gestor do conhecimento}\};$
- $F^I \subset F = \{\text{regras a que não correspondem regras semelhantes em } F_4\};$

Aos conjuntos F , F_4 e F_5 correspondem, respectivamente, os ficheiros *BCCC*, *temp₃* e *temp₄* (Figura 5.8).

A "Assimilação" é o último procedimento no processo desenvolvido pelo GCCC; consiste na assimilação do novo conhecimento para a BCCC, i.e., no seu registo na BCCC como evolução duma regra já existente ou como uma nova regra e corresponde à obtenção do conjunto F :

$$F = F_4 \cup F_5$$

No fim desta fase, o processo de assimilação do conhecimento é considerado completo e a situação é caracterizada por:

- o conhecimento resultante da última actuação da ferramenta de DM está totalmente integrado no conhecimento existente;
- das regras confrontadas, as consideradas sem interesse foram eliminadas;

- de cada nova regra ou de cada nova ocorrência das regras actualizadas fazem já parte:
 - ▶ uma dimensão temporal;
 - ▶ um comentário resultante da participação do gestor do conhecimento e que traduz conhecimento do negócio;
 - ▶ os parâmetros que permitem avaliar da incerteza da regra nesse momento.

Relativamente às quatro fases apresentadas, é possível observar-se que:

- de acordo com o conceito introduzido nesta tese, "**Gestão do CCC**", estão incluídos no GCCC todos os procedimentos necessários que vão desde a obtenção dos padrões de DM até à sua integração na BCCC e que são:
 - ▶ a importação das regras obtidas com DM;
 - ▶ a sua transformação em regras canónicas (Figura 5.4);
 - ▶ o confronto com o conteúdo da BCCC, a apresentação das regras ao gestor do conhecimento organizacional para que este introduza os seus comentários e a actualização da BCCC (Figuras 5.5, 5.6 e 5.8);
- na execução dos procedimentos do GCCC, há uma participação do gestor do conhecimento quer com a introdução de comentários pertinentes para justificar a alteração dos parâmetros observados em cada regra quer com a ordem de eliminação de uma regra considerada sem interesse.

Nos procedimentos desenvolvidos, o registo das regras é efectuado tendo em conta que:

- no mesmo ficheiro, todas as regras registadas têm o mesmo formato;
- o ZACCAR utiliza apenas dois formatos de registo de regras:
 - ▶ o correspondente às regras canónicas e que é utilizado no ficheiro *temp1*;
 - ▶ o correspondente às regras canónicas estendidas é utilizado em todas as outras situações.

Formalmente, os processos presentes no GCCC podem ser sumariados pela Figura 5.9 através das funções matemáticas definidas para a sua materialização.

Na Tabela 5.5 é apresentada a correspondência entre os conjuntos de regras e os ficheiros que os materializam.

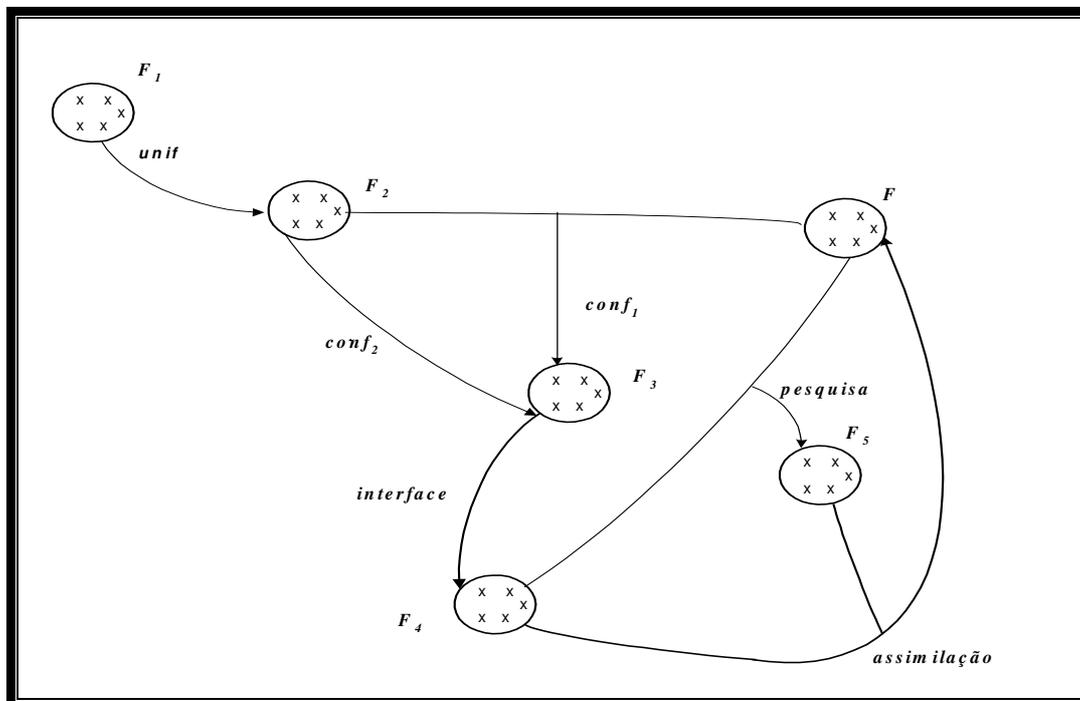


Figura 5.9 – Os processos envolvidos no GCCC.

Tabela 5.5 - Correspondência entre os conjuntos de regras e os ficheiros que os materializam.

Conjuntos	Conteúdo	Ficheiros	Descrição
F	Regras canónicas estendidas que constituem a base de conhecimento.	BCCC	Esta base de conhecimento é uma dos módulos do sistema e é o único de todos os ficheiros apresentados o que permanece após cada actuação do ZACCAR. Em cada sessão, é reconstruído a partir da assimilação do conhecimento existente nos ficheiros <i>temp₃</i> e <i>temp₄</i> .
<i>F₁</i>	Regras obtidas com a ferramenta de DM.	<i>padroes_dm</i>	Ficheiro onde se encontram registados os padrões de comportamento dos clientes, obtidos com uma ferramenta de DM.
<i>F₂</i>	Regras canónicas.	<i>temp₁</i>	Ficheiro onde é registado o conjunto imagem dos padrões existentes em <i>padroes_dm</i> , gerado pela actuação, sobre estes padrões, do procedimento "Uniformização de formatos" materializado pela função <i>unif</i> .
<i>F₃</i>	Regras canónicas estendidas actualizadas no momento presente.	<i>temp₂</i>	Ficheiro onde são registadas as regras resultantes da actuação do procedimento "Confronto", materializado pelas funções <i>conf₁</i> e <i>conf₂</i> .
<i>F₄</i>	Regras canónicas estendidas incluindo comentário do gestor do conhecimento.	<i>temp₃</i>	Ficheiro para onde são transferidas as regras de <i>temp₂</i> que não foram eliminadas pelo gestor do conhecimento, já acrescentadas com os comentários que, até este momento, estavam em falta.
<i>F₅</i>	Regras canónicas estendidas não actualizadas no momento presente.	<i>temp₄</i>	Ficheiro para onde são transferidas as regras da BCCC que não foram actualizadas na sessão presente; este procedimento é materializado pela função <i>pesquisa</i> .

A actividade do GCCC deve ser executada regularmente, provavelmente, com a mesma periodicidade com que actua a ferramenta de DM.

Factores que podem influenciar a execução esporádica desta actividade são, por exemplo, o lançamento de um novo produto, a análise dos resultados de uma campanha de *marketing* ou a realização de um estudo sobre a situação da organização.

5.2.2 A Base de Conhecimento do Comportamento dos Clientes

A BCCC é composta por regras que são sempre registadas na forma de regras canónicas estendidas; esta opção leva a que não haja a necessidade de alterar os módulos do sistema desenvolvido, qualquer que seja a área que forneça padrões através de uma mesma ferramenta (ou algoritmo) de DM. Diferentes ferramentas de DM poderão levar à necessidade da geração de um novo compilador, mantendo-se os restantes módulos do sistema inalterados.

A BCCC deve ter um conteúdo tal que seja elementar a sua integração em qualquer outro sistema de conhecimento da organização ou a sua utilização por aplicações que necessitem de utilizar este tipo de conhecimento. A seguir apresenta-se um exemplo de uma regra da BCCC:

```
regra se estabilidade_emprego='fraca' e saldo_médio < '300 contos'  
então  
opção='não conceder empréstimo'  
evolução  
[(data(01/01/99),comentário("Adesão ao Euro"),parâmetros(n1= 60, n2=70)),  
(data(01/01/86),comentário("Integração na UE"), parâmetros( n1= 70, n2=40)),  
(data(02/05/74),comentário("Deu-se o 25 de Abril"), parâmetros(n1= 90,n2=95))]
```

5.2.3 Consulta à BCCC

Na consulta (Figura 5.10), o consumidor do conhecimento (e.g., membro da organização, aplicação informática) apenas pode ver as regras registadas na BCCC que satisfazem um requisito pré-definido.

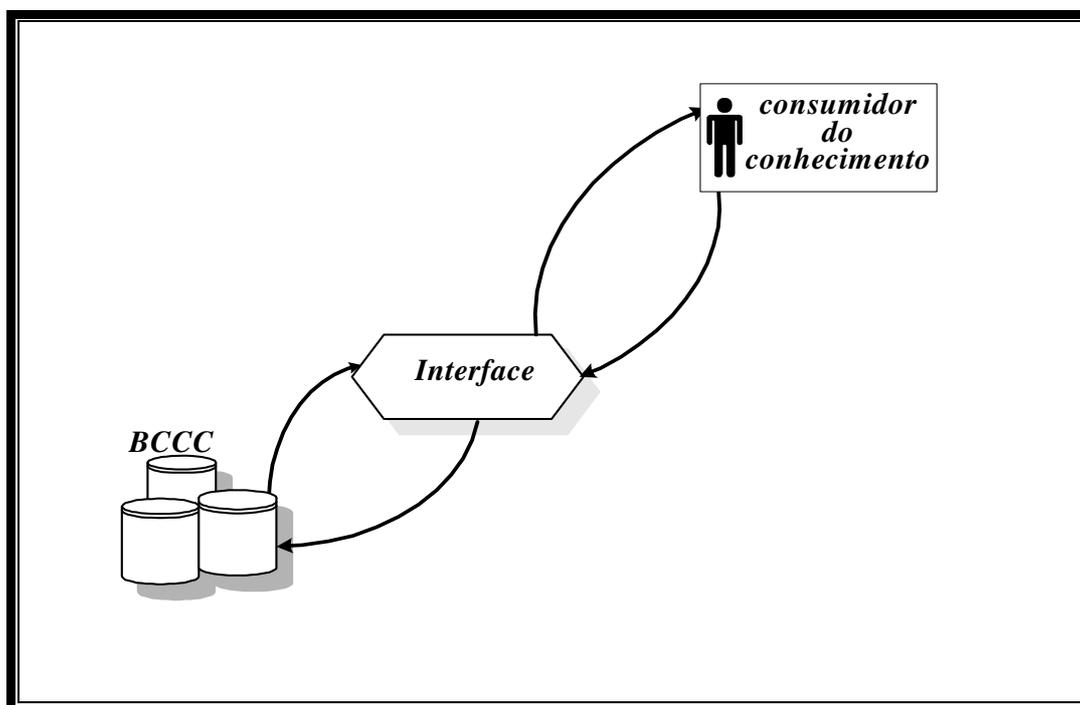


Figura 5.10 – A consulta à BCCC.

Para que se possa efectuar a consulta, desenvolveu-se os seguintes procedimentos:

1. Selecção de atributos

Numa lista de atributos que são mostrados ao consumidor do conhecimento, este selecciona um que entrará no conseqüente das regras a pesquisar.

2. Pesquisa de regras

Tendo em atenção o atributo que foi instanciado na fase de selecção, é feita na BCCC uma pesquisa de todas as regras em que esse atributo entra no conseqüente. Essas regras são visualizadas no écran ou registadas em ficheiro apropriado para utilização posterior.

3. *Interface*

Um *interface* permite que:

- o utilizador visualize os atributos componentes das regras registadas;
- seleccione um atributo da lista;
- visualize ou registre as correspondentes regras.

5.3 Sumário

Este capítulo apresenta a arquitectura, funcionamento e especificação formal do sistema ZACCAR, um sistema de conhecimento cujo objectivo é viabilizar a actividade de zelar pelo CCC. São apresentados os seus módulos constituintes (BCCC, GCCC e consulta) e um conjunto de requisitos a que cada módulo deve obedecer.

Em seguida, é apresentado um protótipo do sistema, tendo em vista demonstrar a viabilidade do conceito proposto e mostrar que é possível criar um sistema computacional que permite que o conhecimento dos padrões dos clientes e a sua evolução sejam mantidos, geridos e integrados, numa forma automática, no conhecimento organizacional após um processo de consolidação do conhecimento existente.

A arquitectura deste sistema permite que ele possa ser encarado não como um sistema fechado mas como um conjunto de componentes com a possibilidade de virem a ser integradas em sistemas mais abrangentes.

A descrição, no capítulo seguinte do estudo de dois casos em que o protótipo foi utilizado tem em vista não só demonstrar a viabilidade do sistema desenvolvido e provar a sua eficácia mas também abrir novas perspectivas para o seu melhoramento.

6. Viabilidade do Sistema ZACCAR

“Prototypes - Working when my friends use them (though they may have to seek my help.)”
(David J. Hand, [Hand, 1994])

Uma actividade organizacional que consiste em zelar pelo CCC, foi apresentada e caracterizada ao longo desta tese. Com o fim de tornar exequível esta actividade, recorrendo a tecnologias existentes, propusemos um sistema – o sistema ZACCAR – descrevemos a sua arquitectura e apresentámos um protótipo. Tal desenvolvimento permitiu a experimentação do sistema, testando a sua funcionalidade e alertando para mudanças necessárias e, mais importante ainda, demonstrou que os objectivos de automação pretendidos podem ser atingidos.

Neste capítulo leva-se a cabo uma investigação laboratorial da dimensão e graus de “sucesso” alcançados bem como se analisam as dificuldades detectadas na sua implementação. Tal foi conseguido através do estudo de dois casos⁸⁴ em áreas distintas de aplicação: turismo e indústria de confecção.

A opção por dois casos e não apenas um poderá permitir o estabelecimento de limitações a possíveis generalizações que, inicialmente, se seria levado a considerar.

Os dois estudos de casos levados a efeito tiveram em vista:

- demonstrar a viabilidade do sistema proposto e a sua aplicabilidade em domínios distintos;

⁸⁴ Um *caso* é um exemplo dum problema que inclui a informação dada, os métodos usados e os resultados obtidos [Stefix, 1995].

- avaliar a sua eficácia;
- identificar pontos fracos do sistema por forma a fornecer pistas para o desenvolvimento de novas opções e melhoramento das já existentes.

Para os casos considerados serão abordados os seguintes pontos:

- descrição do caso (nome, contexto e história, aplicações no negócio e infra-estruturas tecnológicas);
- descrição dos dados (tabelas, atributos, registos e estatísticas);
- utilização de uma ferramenta de DM para a obtenção de padrões de comportamento dos clientes;
- aplicação do protótipo desenvolvido;
- análise crítica dos resultados alcançados;
- análise das dificuldades encontradas naquela utilização e formas de as ultrapassar.

A ferramenta de DM utilizada foi o *Clementine* da SPSS [*Clementine*, 1997]⁸⁵. A opção pelo *Clementine* resultou do facto de que esta ferramenta estava disponível e possuía as características pretendidas:

- algoritmos para gerar regras de produção;
- opção de registo, num ficheiro de texto, das regras obtidas.

Nos dois casos o estudo foi feito ao longo de vários momentos, correspondendo um momento às seguintes situações:

- uma actividade da empresa (e.g., lançamento de novo produto, campanha de *marketing*);
- uma recolha de dados;
- actuação da ferramenta de DM sobre os novos dados;
- actuação do ZACCAR nas novas regras e sua consolidação com o conhecimento já existente na BCCC.

⁸⁵ www.spss.com/Clementine

No momento 1, pelo facto de ainda não existirem regras na BCCC, o ZACCAR limita-se a registar na base de conhecimento as regras obtidas depois da sua transformação em regras canónicas estendidas.

Pelo facto do *Clementine* actuar só sobre os novos dados, a evolução da regra, em cada momento, reflecte, apenas, a situação correspondente a esses dados.

6.1 Caso 1 - Turismo

Neste estudo de caso, o sistema ZACCAR foi aplicado aos dados coleccionados a partir de campanhas de *marketing* sobre uma viagem de fim de curso. Este caso é relacionado com uma sequência de quatro campanhas de *marketing* denominadas por momentos.

Foi efectuado um inquérito (Anexo 4) a alunos de vários cursos de três estabelecimentos de ensino superior, simulando uma situação real: a oferta de uma viagem com destino a Cuba.

O inquérito foi realizado em quatro momentos diferentes a alunos distintos e em número diferente e os dados foram registados numa base de dados com uma estrutura convenientemente definida (Tabela 6.1) (Anexo 5).

As ocorrências registadas na base de dados foram usadas para gerar regras de comportamento (padrões) através do algoritmo GRI (*Generalised Rule Induction*) [*Clementine*, 1997], uma das opções do *Clementine*, utilizando a *stream* apresentada na Figura 6.1.

Os atributos e os seus domínios são descritos na Tabela 6.2.

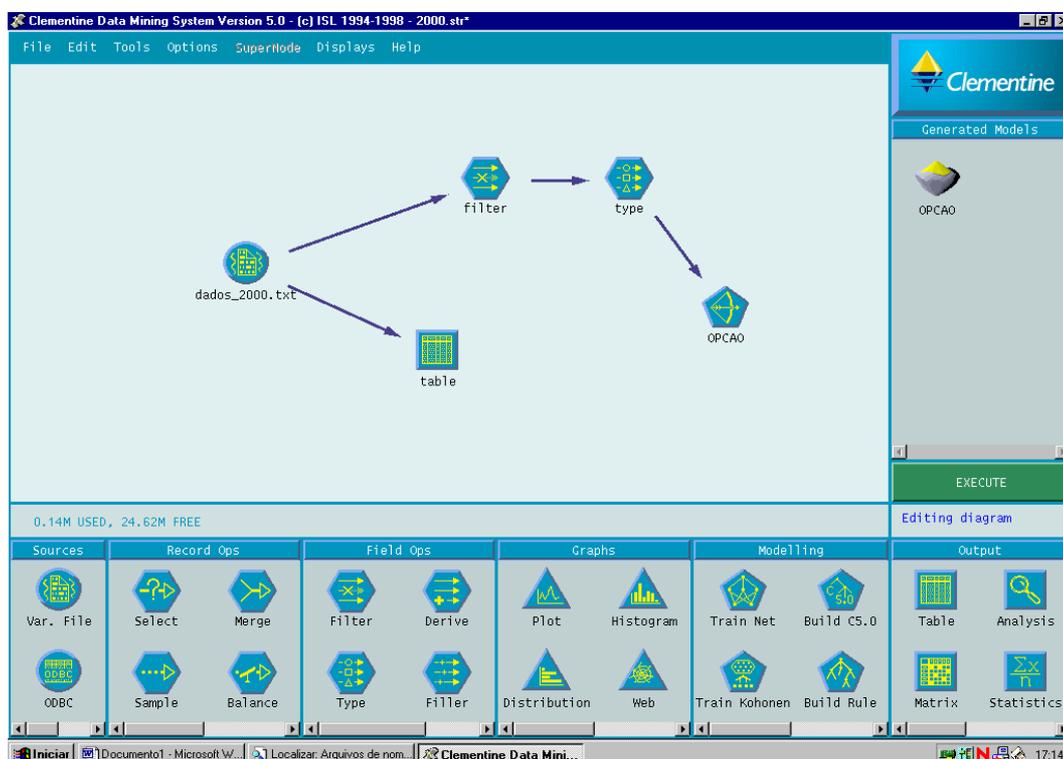


Figura 6.1 – Stream implementada no Clementine no primeiro caso.

Tabela 6.1 - Um subconjunto dos dados das campanhas de marketing.

Situação	Curso	Ano	Idade	Sexo	Tipo_residência	Área_residência	Agregado_familiar	Profissão_pai	Profissão_mãe	Opção
E	5	3	24	F	U	N	3	01	01	S
T	2	2	34	M	U	S	1	01	04	N
T	3	3	30	F	R	C	4	03	04	N
E	1	3	19	F	R	C	5	14	08	S
E	4	4	23	M	R	N	2	12	07	S

Tabela 6.2 - Atributos da base de dados.

Atributo	Descrição do atributo	Domínio	Descrição dos valores do domínio
Situação	A situação do estudante no curso	{E,T}	E = Estudante T = Trabalhador estudante
Curso	Os cursos alvo desta campanha	[1...5]	1 – Contabilidade e Auditoria (ISCAC ⁸⁶) 2 – Gestão (ISCAC) 3 – Informática de Gestão (ISCAC) 4 – Engenharia Civil (UC ⁸⁷) 5 – Informática de Gestão (UM ⁸⁸)

⁸⁶ Instituto Superior de Contabilidade e Administração de Coimbra

⁸⁷ Universidade de Coimbra

⁸⁸ Universidade do Minho

<i>Ano</i>	Ano frequentado no respectivo curso	[1...5]	
<i>Idade</i>	Idade	Inteiro positivo	
<i>Sexo</i>	Sexo	{F,M}	F = Feminino M = Masculino
<i>Tipo_residência</i>	Tipo de residência de origem	{R, U}	R = Rural U = Urbana
<i>Área_residência</i>	Área de residência de origem	{N,C,S,O}	N = Norte C = Centro S = Sul O = Outras (Açores, Madeira, Cabo Verde)
<i>Agregado_familiar</i>	Número de elementos do agregado familiar	N_0 ⁸⁹	
<i>Profissão_pai</i>	Situação profissional do pai	{0,1,...,14,--}	"01" – Agricultor "02" – Comerciante "03" – Dirigente/gestor de empresa "04" – Dona(o) de casa "05" – Estudante "06" – Funcionário administrativo "07" – Empregado comercial "08" – Operário "09" – Professor "10" – Profissional liberal "11" – Quadro médio de empresa/função pública "12" – Quadro superior de empresa/função pública "13" – Técnico especializado "14" – Outros "- -" – Não responde
<i>Profissão_mãe</i>	Situação profissional da mãe	{0,1,...,14,--}	Igual à anterior
<i>Opção</i>	Intenção de realização da viagem proposta	{S, N}	S = Sim N = Não

No estudo deste caso, foram tidos em atenção os seguintes pontos:

- o atributo *Opção*, variável a estudar, é o único atributo de saída ao longo dos quatro momentos;
- regras idênticas são as regras em que os conjuntos dos pares (atributo, valor) são idênticos, independentemente da ordem em que se apresentam.

⁸⁹ Conjunto dos números naturais, incluindo o zero.

Em cada momento, a ferramenta de DM actuou apenas sobre os dados obtidos nesse momento, produzindo as regras e, para cada uma, os respectivos níveis de suporte e de confiança.

Das regras obtidas (Anexo 6), seleccionou-se duas que foram seguidas para exemplificar o comportamento do sistema ao longo dos quatro momentos.

As regras seleccionadas para exemplificar o estudo foram detectadas nos quatro momentos do estudo; os seus graus de confiança bem como os das outras regras obtidas são de valor muito reduzido provavelmente pela população em estudo ser muito heterogénea.

Houve a preocupação de seleccionar estas duas regras entre as que apresentam como consequente *Opção = "S"* o que corresponde a padrões de comportamento de potenciais clientes das viagens oferecidas.

6.1.1. Momento 1

No momento 1, a proposta é uma viagem a Cuba, com uma permanência de 8 dias/7 noites na praia de Varadero, instalações em hotel de 4 estrelas em regime de TI (Tudo Incluído) e com um preço por pessoa em quarto duplo de 900 euros.

Foram inquiridos 510 alunos e a distribuição, para cada um dos atributos dos dados recolhidos após esta campanha está representada na Tabela 6.3.

Tabela 6.3 - A distribuição de cada atributo no momento 1.

Situação	<i>Estudante</i>		<i>Trabalhador- estudante</i>		
	62%		38%		
Curso	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
	68%	20%	3%	1%	8%
Ano	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
	18%	31%	40%	10,7%	0,3%
Idade	<i>18-20</i>	<i>21-27</i>	<i>28-35</i>	<i>>35</i>	
	15%	70%	11%	4%	
Sexo	<i>Masculino</i>		<i>Feminino</i>		

	37%	63%														
Tipo_residência	<i>Rural</i>	<i>Urbana</i>														
	67%	33%														
Área_residência	<i>Norte</i>	<i>Centro</i>	<i>Sul</i>	<i>Outros</i>												
	12%	83%	4%	1%												
Agregado_familiar	<i><=2</i>	<i>3-5</i>	<i>>5</i>													
	16%	82%	2%													
	--	01	02	03	44	05	06	07	08	09	10	11	12	13	14	
Profissão_pai	6%	6%	11%	9%	0%	0%	3%	4%	11%	1%	4%	8%	2%	3%	32%	
Profissão_mãe	3%	2%	8%	3%	39%	0%	6%	3%	6%	4%	0%	6%	2%	0%	18%	
Opção	<i>Sim</i>	<i>Não</i>														
	30%	70%														

Os padrões obtidos com o *Clementine* (Anexo 6.1) foram:

Regra 1:

*Opcao == S <= Situação == E & Curso == 5 & Idade > 22.5 (15:2.941%, 0.73)*⁹⁰

Regra 2:

Opcao == S <= Idade < 19.5 & Área de residência == C & Agregado familiar > 3.5 (23:4.51%, 0.35)

O ZACCAR, através do procedimento Uniformização de Formatos, transformou estas regras nas respectivas regras canônicas:

Regra 1:

*regra se situacao='E' e curso='5' e idade>22.5
entao
opcao = 'S'
parametros
[n1=2.941,n2= 73].*

Regra 2:

*regra se idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5
entao
opcao = 'S'
parametros
[n1=4.51,n2= 35].*

⁹⁰ Estes valores representam, respectivamente, o número de ocorrências na base de dados que verificam os antecedentes da regra, o nível de suporte em porcentagem e o nível de confiança num intervalo [0,1].

Como este é o momento inicial, o ZACCAR transformou estas regras em regras canónicas estendidas e registou-as, após a inclusão de comentários do gestor do conhecimento, na BCCC na forma:

Regra 1:

*regra se situação='E' e curso='5' e idade>22.5
entao
opção = 'S'
evolucao
[(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'), parâmetros (n1=2.941,
n2=73))].*

Regra 2:

*regra se idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5
entao
opção = 'S'
evolucao
[(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'),
parâmetros(n1=4.51,n2=35))].*

Na evolução das regras pode-se ver:

- a data de registo das regras (30/10/2000);
- um comentário sobre a campanha conducente a estas regras ("Viagem proposta - Cuba (Varadero));
- os valores dos parâmetros níveis de suporte e confiança das regras (e.g., $n_1=4,51$, $n_2=35$).

Como se pode observar, esta campanha teve uma maior aceitação por parte dos estudantes que se podem englobar no perfil traduzido pela regra 1.

6.1.2. Momento 2

No momento 2, a campanha de *marketing* propunha uma viagem a Cuba com uma estadia de 7 dias/6 noites em Varadero e 1 dia/1 noite em Havana, sendo as outras características idênticas às da viagem anterior.

Foram inquiridos 357 alunos e a distribuição em cada atributo, no momento 2, está representada na Tabela 6.4.

Tabela 6.4 - A distribuição de cada atributo no momento 2.

Situação	<i>Estudante</i>		<i>Trabalhador-estudante</i>												
	60%		40%												
Curso	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>										
	71%	20%	2%	1%	6%										
Ano	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>										
	17%	31%	42%	9%	1%										
Idade	<i>18-20</i>	<i>21-27</i>	<i>28-35</i>	<i>>35</i>											
	14%	72%	10%	4%											
Sexo	<i>Masculino</i>					<i>Feminino</i>									
	37%					63%									
Tipo_residência	<i>Rural</i>					<i>Urbana</i>									
	67%					33%									
Área_residência	<i>Norte</i>	<i>Centro</i>	<i>Sul</i>	<i>Outros</i>											
	11%	85%	3%	1%											
Agregado_familiar	<i><=2</i>	<i>3-5</i>	<i>>5</i>												
	16%	82%	2%												
Profissão_pai	<i>--</i>	<i>01</i>	<i>02</i>	<i>03</i>	<i>04</i>	<i>05</i>	<i>06</i>	<i>07</i>	<i>08</i>	<i>09</i>	<i>10</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>
	8%	6%	10%	7%	0%	0%	3%	4%	11%	1%	4%	9%	2%	3%	32%
Profissão_mãe	<i>3%</i>	<i>2%</i>	<i>8%</i>	<i>3%</i>	<i>38%</i>	<i>0%</i>	<i>6%</i>	<i>4%</i>	<i>5%</i>	<i>4%</i>	<i>1%</i>	<i>6%</i>	<i>1%</i>	<i>0%</i>	<i>19%</i>
Opção	<i>Sim</i>					<i>Não</i>									
	32%					68%									

Neste momento, os padrões obtidos com o *Clementine* (Anexo 6.2) foram:

Regra 1:

$Opcao == S \leq Situa\c{c}ao == E \ \& \ Curso == 5 \ \& \ Idade > 22.5$ (4:1%, 0.50)

Regra 2:

$Opcao == S \leq Idade < 19.5 \ \& \ Area_residencia == C \ \& \ Agregado_familiar > 3.5$ (15:4.202%, 0.73)

Após uniformização com o ZACCAR, as respectivas regras canônicas são:

Regra 1:

regra se *situacao='E' e curso='5' e idade>22.5*
entao
opcao = 'S'
parametros
[n1=1,n2= 50].

Regra 2:

regra se *idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5*
entao
opcao = 'S'
parametros
[n1=4.202,n2= 73].

Havendo já uma BCCC criada no momento anterior, o passo seguinte consistiu no confronto das regras canónicas estendidas já existentes com estas novas regras, actualizando aquelas e registando-as na BCCC com a forma:

Regra 1:

regra se *situacao='E' e curso='5' e idade>22.5*
entao
opcao = 'S'
evolucao
*[(data (30,11,2000), comentario('Depois da campanha de marketing 1 – Varadero+Havana'),
parâmetros (n1=1, n2=50)),
(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'),
parâmetros(n1=2.941,n2=73))].*

Regra 2:

regra se *idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5*
entao
opcao = 'S'
evolucao
*[(data (30,11,2000), comentario('Depois da campanha de marketing 1 – Varadero+Havana'),
parâmetros (n1=4.202, n2=73)),
(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'),
parâmetros(n1=4.51,n2=35))].*

Após a actualização, a evolução de cada regra apresenta, além do anteriormente existente, a nova data de actualização (30/11/2000), um comentário sobre a campanha levada a cabo e os respectivos parâmetros (níveis de suporte e confiança).

Como se pode observar pela consulta destas regras, os estudantes que satisfazem o perfil traduzido na regra 1, deram preferência à primeira campanha ($n_2=73\%$ na

campanha 1 e $n_2=50\%$ na campanha 2) enquanto que os que se podem incluir num perfil correspondente à 2ª regra preferiram a 2ª campanha ($n_2=35\%$ na campanha 1 e $n_2=73\%$ na campanha 2).

6.1.3. Momento 3

A campanha de *marketing* lançada no momento 3 propõe uma viagem a Cuba com a estadia de 7 dias/6 noites em Varadero, o pagamento feito em 3 prestações e as outras características idênticas às da campanha 1.

Foram inquiridos 243 alunos e a distribuição por cada atributo é a que se representa na Tabela 6.5.

Tabela 6.5 - A distribuição de cada atributo no momento 3.

Situação	<i>Estudante</i>	<i>Trabalhador- estudante</i>													
	58%	42%													
Curso	1	2	3	4	5										
	72%	19%	3%	1%	5%										
Ano	1	2	3	4	5										
	19%	31%	41%	9%	0%										
Idade	18-20	21-27	28-35	>35											
	12%	74%	9%	5%											
Sexo	<i>Masculino</i>		<i>Feminino</i>												
	36%		64%												
Tipo_residência	<i>Rural</i>	<i>Urbana</i>													
	70%	30%													
Área_residência	<i>Norte</i>	<i>Centro</i>	<i>Sul</i>	<i>Outros</i>											
	11%	85%	3%	1%											
Agregado_familiar	<=2	3-5	>5												
	17%	80%	3%												
Profissão_pai	--	01	02	03	04	05	06	07	08	09	10	11	12	13	14
	10%	7%	7%	7%	0.4%	0%	2.3%	5%	10%	1%	2%	8%	2%	2.3%	36%
Profissão_mãe	4%	3%	7%	2%	37%	0%	6%	3%	6%	3%	0%	7%	1%	0%	21%
Opção	<i>Sim</i>		<i>Não</i>												
	17%		83%												

Neste momento, com a actuação do *Clementine*, obteve-se os padrões (Anexo 6.3):

Regra 1:

Opção == S <= Situação == E & Curso == 5 & Idade > 22.5 (4:1.65%, 0.50)

Regra 2:

Opção == S <= Idade < 19.5 & Area residencia == C & Agregado familiar > 3.5 (4:1,65%, 0.25)

O procedimento Uniformização de Formatos transformou estas regras nas regras canónicas:

Regra 1:

*regra se situação='E' e curso='5' e idade>22.5
entao
opção = 'S'
parametros
[n1=1.65 ,n2= 50].*

Regra 2:

*regra se idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5
entao
opção = 'S'
parametros
[n1=1.65,n2= 25].*

O confronto entre o conteúdo da BCCC e estas regras conduziu a:

Regra 1:

*regra se situação='E' e curso='5' e idade>22.5
entao
opção = 'S'
evolucao
[(data (30,12,2000), comentario('Depois da campanha de marketing 2 – Varadero, pago a prestações'), parametros (n1=1.65, n2=50)),
(data (30,11,2000), comentario('Depois da campanha de marketing 1 – Varadero+Havana'),
parametros (n1=1, n2=50)),
(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'),
parametros(n1=2.941,n2=73))].*

Regra 2:

*regra se idade<19.5 e area_residencia = 'C' e agregado_familiar > 3.5
entao
opção = 'S'*

evolucao

[(data (30,12,2000), comentario(`Depois da campanha de marketing 2 – Varadero, pago a prestações `), parâmetros (n1=1.65, n2=25)),
 (data (30,11,2000), comentario(`Depois da campanha de marketing 1 – Varadero+Havana`), parâmetros (n1=4.202, n2=73)),
 (data (30,10,2000), comentario(`Viagem proposta – Cuba (Varadero)`), parâmetros (n1=4.51, n2=35))].

Após a actualização, a evolução de cada regra regista a informação correspondente a este terceiro momento, i.e., a data da actualização (30/12/2000), um comentário sobre a campanha acabada de levar a efeito e os correspondentes parâmetros (níveis de suporte e de confiança).

Pela consulta destas regras, registadas na BCCC, o gestor do negócio pode concluir que os estudantes cujo perfil é descrito por estas regras.

6.1.4. Momento 4

No momento 4, a campanha de *marketing* propõe uma viagem de 6 dias/5 noites a Varadero, 1 dia/1 noite em Havana, o pagamento da viagem em 3 prestações, sendo as outras características idênticas às correspondentes nas opções anteriores.

Foram inquiridos 201 alunos e a distribuição para cada atributo é a que se mostra na Tabela 6.6.

Tabela 6.6 - A distribuição de cada atributo no momento 4.

Situação	Estudante			Trabalhador- estudante	
	57%			43%	
Curso	1	2	3	4	5
	73%	18%	3%	1%	5%
Ano	1	2	3	4	5
	20%	32%	39%	9%	0%
Idade	18-20	21-27	28-35	>35	
	11%	73%	10%	6%	
Sexo	Masculino			Feminino	
	35%			65%	

Tipo_residência	<i>Rural</i>	<i>Urbano</i>														
	70%	30%														
Área_residência	<i>Norte</i>	<i>Centro</i>	<i>Sul</i>	<i>Outros</i>												
	10%	85%	3%	2%												
Agregado_familiar	<i><=2</i>	<i>3-5</i>	<i>>5</i>													
	17%	81%	2%													
	--	01	02	03	04	05	06	07	08	09	10	11	12	13	14	
Profissão_pai	11%	8%	7%	6%	0%	0%	2%	4%	11%	0%	2%	8%	2%	3%	36%	
Profissão_mãe	5%	2%	8%	2%	39%	0%	5%	2%	6%	3%	0%	6%	1%	0%	21%	
Opção	<i>Sim</i>		<i>Não</i>													
	53%		47%													

Neste momento, o padrão correspondente à regra 1 não foi detectado. O padrão correspondente à regra 2, obtido com o *Clementine* (Anexo 6.4) foi:

Regra 2:

Opcao == S <= Idade < 19.5 & Area residencia == C & Agregado familiar > 3.5 (3:1,49%, 0.67)

A correspondente regra canónica toma a forma:

Regra 2:

*regra se idade<19.5 e area_residencia = `C` e agregado_familiar > 3.5
entao
opcao = `S`
parametros
[n1=1.49,n2= 67].*

O confronto desta regra com o conteúdo da BCCC conduz à regra canónica estendida já incluindo esta actualização na forma:

Regra 2:

*regra se idade<19.5 e area_residencia = `C` e agregado_familiar > 3.5
entao
opcao = `S`
evolucao
[(data (30,01,2001), comentario(`Depois da campanha de marketing 3 – Varadero+Havana, pago a prestações `), parametros (n1=1.49, n2=67)),
(data (30,12,2000), comentario(`Depois da campanha de marketing 2 – Varadero, pago a prestações `), parametros (n1=1.65, n2=25)),
(data (30,11,2000), comentario(`Depois da campanha de marketing 1 – Varadero+Havana`), parametros (n1=4.202, n2=73)),*

*(data (30,10,2000),comentario(`Viagem proposta – Cuba (Varadero)`),
parâmetros($n_1=4.51,n_2=35$))].*

Aqui, a evolução apresenta os quatro momentos correspondentes às quatro campanhas realizadas, podendo-se observar a data de actualização, um comentário sobre a campanha levada a cabo no momento 4 e os respectivos parâmetros (níveis de suporte e de confiança).

Uma consulta à BCCC, neste momento, permite-nos tirar algumas conclusões sobre o comportamento das pessoas alvo das campanhas, como sejam:

- Na regra 1:

Indivíduos com o perfil representado nesta regra i.e., com situação de estudante, pertencentes ao curso ao qual foi atribuído o número 5 e com idade superior a 22.5 anos, deram preferência à campanha 1 ($n_2=73$) sobre a campanha 2 ($n_2=50$); logo, seria de esperar que dessem preferência à campanha 3 relativamente à campanha 4 o que, na realidade, é verdade pois n_2 toma o valor de 50% na 3ª campanha nem sequer tendo sido detectado o padrão na 4ª campanha.

- Na regra 2:

Os indivíduos com o perfil representado nesta regra, i.e., com idade inferior a 19,5 anos, área de residência no Centro e com agregado familiar com mais do que 3 elementos, deram preferência à 2ª campanha relativamente à 1ª, logo seria de esperar que dessem preferência à 4ª campanha relativamente à 3ª o que se verifica na realidade, pois o nível de confiança é, respectivamente, 67% e 25%.

Consultando as regras 1 e 2 na sua forma canónica estendida, pode-se concluir que as campanhas que consideram o pagamento faseado têm menor aceitação do que as correspondentes campanhas com pagamento efectuado por uma única vez o que nos permite concluir que os estudantes inquiridos valorizam muito mais o tempo de estadia do que as condições de pagamento.

6.1.5. Padrões de comportamento correspondentes a reacções negativas às campanhas

Idealmente, um inquirido aceita a oferta proposta num dos quatro momentos da campanha. No entanto, tal não aconteceu com alguns grupos de estudantes.

A análise às reacções negativas às campanhas pode-se revelar muito útil na medida em que se podem obter elementos sobre potenciais clientes onde o tipo de *marketing* desenvolvido não é aconselhável e para os quais há necessidade de pensar em novo tipo de campanhas.

Algumas regras deste tipo foram registadas na base de conhecimento construída como, por exemplo:

```
regra se curso='2' e ano>'1' e profissao_pai='14'  
entao  
opcao = 'N'  
evolucao  
[(data (30,01,2001), comentario('Depois da campanha de marketing 3 – Varadero+Havana, pago a prestações '), parâmetros (n1=5, n2=50)),  
(data (30,12,2000), comentario('Depois da campanha de marketing 2 – Varadero pago a prestações '), parâmetros (n1=5, n2=77)),  
(data (30,11,2000), comentario('Depois da campanha de marketing 1 – Varadero+Havana'), parâmetros (n1=4, n2=100)),  
(data (30,10,2000),comentario('Viagem proposta – Cuba (Varadero)'), parâmetros(n1=3,n2=81))].
```

Como se pode observar nesta regra, um número elevado de alunos do curso de Gestão do ISCAC que já não se encontram matriculados no primeiro ano e cujos pais têm profissões indiferenciadas (Tabela 6.2) não realizariam a viagem proposta em nenhum dos quatro momentos; no estudo, ao analisarmos a evolução da regra, apercebem-nos de que essa opção é mais elevada no momento 2 ($n_2=100$), diminuindo até $n_2=50$ no momento 4.

Num estudo mais aprofundado seria vantajoso, por exemplo, registar, nas bases de dados, outras informações não tratadas neste estudo como sejam os motivos que levaram alguns estudantes a rejeitar todas as campanhas propostas.

6.1.6. Discussão

Ao longo de um período dividido em quatro momentos, estudámos a evolução de duas regras destacadas da base de conhecimento. A forma canónica estendida das regras e a sua evolução demonstraram a sua utilidade para:

- a análise do impacto que as campanhas de *marketing* tiveram nos estudantes inquiridos;
- a orientação para novas actividades de CRM, nomeadamente futuras campanhas de *marketing* ou a repetição, em segmentos específicos da população estudantil, das campanhas 1 e 2;
- a detecção de diferenças significativas entre os parâmetros de uma mesma regra em diferentes momentos e a possibilidade de analisar a evolução temporal das diferenças.

Após a elaboração deste estudo, duas características importantes do sistema podem ser apontadas:

- a possibilidade de actualizar o CCC de uma forma automática;
- a optimização de circuitos de informação;
- a facilidade de consulta de uma base de conhecimento dinamicamente actualizada onde cada membro da organização pode observar a evolução das diferentes regras de comportamento.

Este primeiro estudo de caso permitiu a detecção de situações não previstas e que carecem de ser analisadas a fim de se encontrar a melhor solução para o seu tratamento como sejam:

- a detecção de novas regras e o abandono de outras no lançamento de novas campanhas;
- a constatação de que o conhecimento descoberto pode ser redundante.

Vejam os exemplos desta última situação, detectada no momento 2:

Regra 1:

regra se profissao_pai="10"

entao

opcao="S"

parametros

[4,71]

e

Regra 2:

regra se curso="1" e profissao_pai="10"

entao

opcao="S"

parametros

[2,88]

Como se pode observar, os estudantes com o perfil definido na segunda regra formam um subconjunto daqueles cujo perfil é o indicado na primeira regra. Neste estudo em análise, consideramos que a primeira regra deve ser eliminada pois futuras campanhas de *marketing* com características idênticas à levada a cabo neste momento, deverão ser orientadas aos estudantes com o perfil caracterizado na segunda regra.

Contudo, o protótipo desenvolvido contempla como regras semelhantes as que possuem iguais pares (atributo, valor) para o antecedente e para o consequente; de acordo com esta medida de semelhança, ambas as regras são registadas na BCCC.

A incorporação no ZACCAR de uma funcionalidade que permita o registo de apenas uma das regras depende, pois, de um aprofundamento da noção de similaridade entre duas regras. Essa funcionalidade, a desenvolver na fase de assimilação, deve permitir, face a uma base de conhecimento do domínio, uma das hipóteses seguintes:

- manter ambas as regras;
- eliminar uma das regras, optando por:
 - ▶ uma generalização (corresponde, no exemplo anterior, a manter a regra 1);
 - ▶ uma particularização (corresponde, no exemplo anterior, a manter a regra 2).

6.2 Caso 2 - Indústria de Confeção

Neste estudo de caso, deu-se particular importância à análise do processo de consolidação do conhecimento descoberto através da utilização de ferramentas de DM, uma funcionalidade chave do ZACCAR.

Os dados utilizados neste caso correspondem aos registos existentes nas bases de dados transaccionais da AMMA⁹¹ (Anexo 7), uma fábrica com cerca de 25 anos de existência, sediada na zona centro de Portugal, que se dedica à confeção de roupa para homem, comercializando uma conhecida marca⁹².

O desconhecimento sobre este negócio e sobre a estrutura das bases de dados da empresa e do seu conteúdo, levou à necessidade de uma fase (Fase 1 na Figura 6.2) que permitisse a aquisição de tal conhecimento. Esta fase foi dividida em três passos - compreensão do negócio, compreensão dos dados e preparação dos dados - seguindo a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*)⁹³ [Rodrigues *et al.*, 2000].

O estudo foi, como o anterior, dividido em momentos os quais correspondem a diferentes campanhas levadas a cabo pela empresa. Em cada momento, foram tidos em consideração os passos envolvidos na Fase 2 da Figura 6.2.

⁹¹ www.ubi.pt/festadamoda/e_conf/amma.html

⁹² Por razões de confidencialidade esta marca não é revelada.

⁹³ www.crisp-dm.org. Consultado em 25/09/2000.

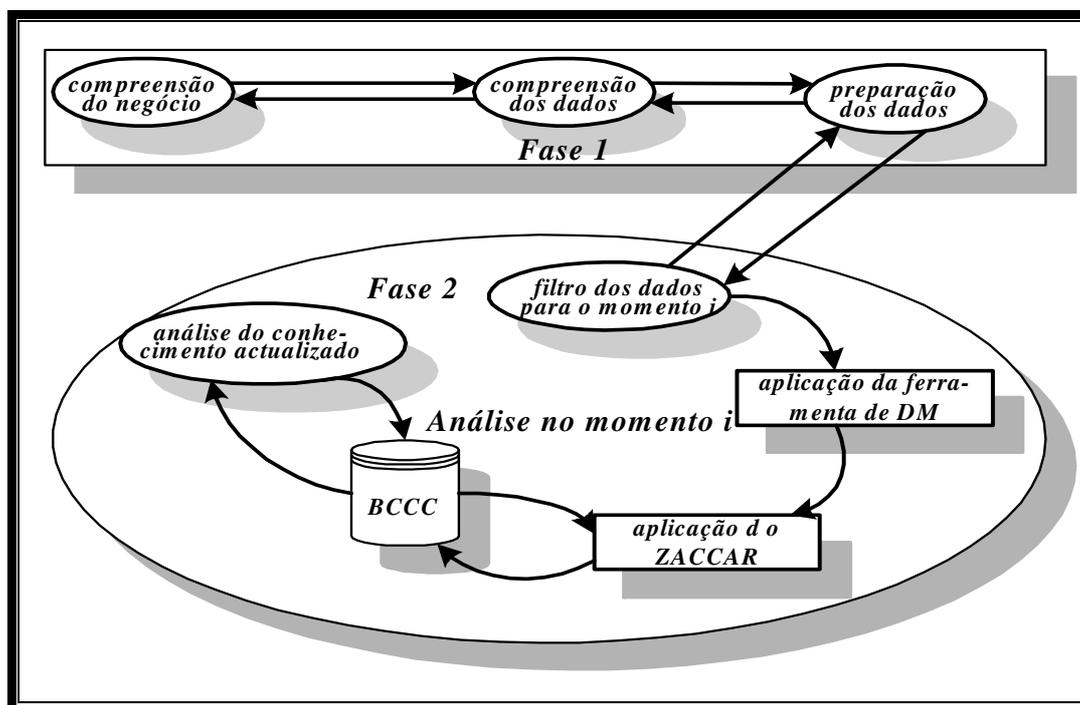


Figura 6.2 – Metodologia seguida no estudo de caso.

6.2.1 Fase 1

6.2.1.1 Compreensão do Negócio

Nesta fase, tentámos, junto do gestor da empresa, compreender o negócio quer nos seus objectivos quer no modo de funcionamento, tendo-nos sido, também, prestadas informações sobre as campanhas de *marketing* levadas a efeito e transmitido algum do conhecimento existente sobre os clientes, resultante do contacto directo cliente-empresa ao longo dos anos.

Inicialmente, as vendas dirigiam-se apenas ao mercado nacional mas, nos últimos anos, atendendo a mudanças entretanto verificadas, houve a necessidade de procurar clientes no mercado internacional o que tem sido conseguido com algum sucesso; actualmente, começam a ter algum peso mercados como o francês, o espanhol e sobretudo o inglês.

As vendas efectuadas são-no apenas para os estabelecimentos comerciais não havendo, salvo muito raras excepções, um contacto directo com o utilizador da roupa confeccionada. Deste modo, o cliente para esta empresa, é o estabelecimento comercial.

Dada a elevada qualidade das confecções e, conseqüentemente, os preços praticados, o mercado alvo é altamente seleccionado.

O negócio é do tipo sazonal, havendo, fundamentalmente, duas entregas anuais:

- em Janeiro/Fevereiro, relativa à colecção de Verão;
- em Julho/Agosto relativa à colecção de Inverno.

Sendo um produto clássico, os artigos comercializados mantêm, ao longo do tempo, um conjunto de características comuns, variando, eventualmente, na qualidade dos tecidos e cores.

Uma colecção é apresentada com cerca dum ano de antecedência relativamente à sua comercialização, em eventos da especialidade na sequência dos quais os clientes fazem as suas encomendas que são satisfeitas nos períodos já definidos.

Tem havido algumas campanhas publicitárias, nomeadamente, no patrocínio de programas na televisão ou na apresentação de faixas em campos de futebol durante jogos importantes.

Não tem sido notado o efeito directo de tais campanhas mas também não existem processos implementados na empresa para fazer tal análise.

Informações do mercado são obtidas, com frequência, através de contactos com vendedores de outras empresas do mesmo ramo.

O conhecimento sobre o cliente é um conhecimento directo e tal tem sido suficiente porque, por enquanto, o número de clientes é reduzido. Com a abertura ao mercado internacional pode ser que, a breve prazo, este conhecimento se torne insuficiente, sendo, pois, necessário, arranjar outros processos para a sua obtenção.

Baseados apenas no seu conhecimento sobre o negócio, os gestores da empresa descreveram um conjunto de regras sobre o comportamento dos seus clientes como, por exemplo:

- nota-se, ao longo dos anos, uma fidelização dos clientes à empresa;

- o mercado nos diversos países tem características diferentes e, conseqüentemente, exigências também diferentes;
- o mercado inglês é aquele que mais se diferencia, exigindo produtos com características muito específicas;
- o mercado brasileiro que nunca foi muito expressivo, tende a desaparecer.

6.2.1.2 Compreensão dos Dados

Nesta fase, houve a preocupação de analisar os dados a explorar, tendo em vista compreender o significado de cada um dos atributos das bases de dados da empresa e definir estratégias para o seu tratamento, selecção e manuseamento.

Inicialmente, a empresa possuía um sistema informático onde registava um conjunto de dados embora a maior parte fosse tratada manualmente. Desde 1998, existe um novo sistema instalado, sendo todos os dados quer de encomendas quer de compras aí registados. Quando a empresa optou por este sistema, não foi feita uma migração dos dados anteriores; dada esta situação, apenas se considera, para este estudo, a informação registada a partir daquela data.

A aplicação instalada, desenvolvida em *Cobol (Common Business Oriented Language)* por uma *software-house* portuguesa, não é específica da AMMA mas sim comum a diversas outras empresas dentro do mesmo ramo; apenas alguma parametrização é feita à medida de cada uma das empresas onde está instalada.

Tendo em vista a exploração dos dados registados nesta base de dados, recorreremos ao técnico da empresa que desenvolveu o *software* o qual se disponibilizou de imediato para nos fornecer todas as informações necessárias.

Entre os dados registados nas bases de dados da empresa⁹⁴, seleccionámos aqueles que considerámos úteis para o presente estudo, ficando reduzidos a quatro tabelas cuja estrutura se apresenta nas Tabelas 6.7, 6.8, 6.9 e 6.10.

⁹⁴ O conteúdo desta base de dados não é disponibilizado em anexo, procedimento seguido no caso anterior, pela necessidade de manter a confidencialidade dos dados da empresa. Assim e apenas para exemplificar, é apresentado no Anexo 6 um extracto daquela base de dados.

Tabela 6.7 - Estrutura da tabela de clientes.

Atributo	Descrição do atributo	Tipo	Dimensão
<i>clinumero</i>	Número do cliente	caracter	9
<i>clidir</i>	Morada do cliente	caracter	35
<i>clinoe</i>	Cidade de residência	caracter	35
<i>clicpe</i>	Código postal	caracter	35
<i>clicsel</i>	País	caracter	5

Nesta tabela existem dados correspondentes a 421 clientes.

Tabela 6.8 - Estrutura da tabela de artigos.

Atributo	Descrição do atributo	Tipo	Dimensão
<i>artmod</i>	Código 1 do modelo	caracter	8
<i>artmod10</i>	Código 2 do modelo	caracter	6
<i>artmodvar</i>	Código 3 do modelo	caracter	2
<i>artcol</i>	Cor do artigo	caracter	10
<i>artdes</i>	Descrição do artigo	caracter	25

Estão referenciados 147 produtos.

Tabela 6.9 - Estrutura da tabela de cabeçalho de encomendas.

Atributo	Descrição do atributo	Tipo	Dimensão
<i>capnum</i>	Número da encomenda	caracter	9
<i>capcli</i>	Número do cliente	caracter	35
<i>caprep</i>	Agente	caracter	35
<i>capult</i>	Condições de pagamento	caracter	35
<i>capfec</i>	Data da encomenda	data	8
<i>capfva</i>	Data da satisfação da encomenda	data	8

Nesta base de dados estão registadas 5953 encomendas podendo, uma mesma encomenda, incluir vários produtos. O número de produtos adquiridos é de 323821.

Tabela 6.10 - Estrutura da tabela de pedidos.

Atributo	Descrição do atributo	Tipo	Dimensão
<i>pednum</i>	Número de encomenda	numérico	6
<i>pedlin</i>	Número de linha na encomenda	numérico	3
<i>pedm01</i>	Número do modelo pedido	numérico	8
<i>pedc01</i>	Número que engloba referência, cor e escala do modelo encomendado	caracter	10
<i>pedcli</i>	Número do cliente	caracter	9

<i>pedpe1i</i>	Quantidade da medida 1	numérico	6
<i>pedpe2i</i>	Quantidade da medida 2	numérico	6
<i>pedpe3i</i>	Quantidade da medida 3	numérico	6
<i>pedpe4i</i>	Quantidade da medida 4	numérico	6
<i>pedpe5i</i>	Quantidade da medida 5	numérico	6
<i>pedpe6i</i>	Quantidade da medida 6	numérico	6
<i>pedpe7i</i>	Quantidade da medida 7	numérico	6
<i>pedpe8i</i>	Quantidade da medida 8	numérico	6
<i>pedpe9i</i>	Quantidade da medida 9	numérico	6
<i>pedpe0i</i>	Quantidade da medida 10	numérico	6
<i>pedpe11i</i>	Quantidade da medida 11	numérico	6
<i>pedpe12i</i>	Quantidade da medida 11	numérico	6

Dos atributos referenciados nas tabelas anteriores, alguns devem ser analisados cuidadosamente pois podem-se decompor ou agregar em informação útil para o estudo a efectuar. Assim, por exemplo, o código de qualquer um dos produtos que a empresa comercializa tem uma dimensão de 8 dígitos e a estrutura que se apresenta na Figura 6.3.

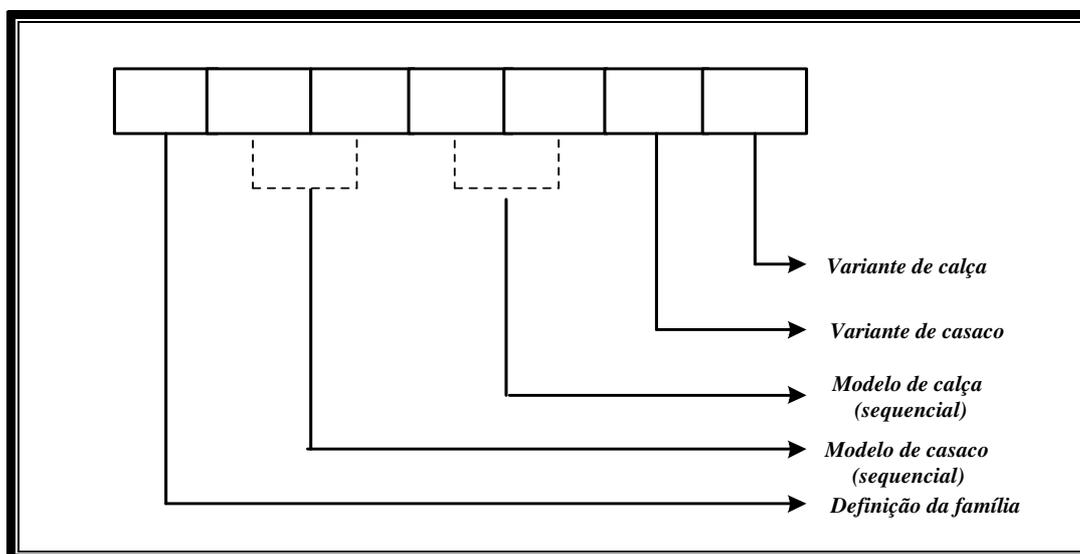


Figura 6.3 – Código do produto.

Os campos indicados no código do produto podem tomar os valores da Tabela 6.11.

Tabela 6.11 - Valores de campos no código dum produto.

	<i>Definição da família</i>		<i>Variante do casaco</i>		<i>Variante da calça</i>
1	Casaco	0	Sem aberturas	0	Sem acabamentos
2	Fato	1	Com 1 abertura	1	Bainha acabada
3	Fato com colete	2	Com 2 aberturas	2	Com virola
4	Calça				
5	Colete				
6	Smoking				
7	Fato de smoking				
8	Fraque/Labita				
9	Sobretudo				

Exemplo: Código do produto “Calça branca com virola”: 4001202.

Esta etapa de compreensão dos dados, permitiu-nos, também, a sua separação nos momentos mais convenientes para este estudo, os quais foram definidos de acordo com as campanhas de *marketing* efectuadas.

6.2.1.3 Preparação dos Dados

Após uma análise dos dados e dos correspondentes atributos já seleccionados entre algumas dezenas registadas nas bases de dados e que não traziam qualquer mais valia para este estudo, seguiu-se uma fase em que, aplicando um conjunto de operadores do SGBD *Foxpro* foi construída uma estrutura de tabela que nos pareceu ser a ideal para a passagem à etapa seguinte: aplicação de uma ferramenta de DM.

As operações executadas no *Foxpro* foram comandos de SQL como, por exemplo, a eliminação de alguns atributos, a obtenção de atributos derivados de atributos já existentes, a limpeza e a correcção de dados (e.g., fazer que em cada campo da base de dados existissem apenas letras minúsculas ou maiúsculas, eliminar espaços em branco no início dos campos).

Os atributos que integram os dados a analisar são: *produto/modelo, data de encomenda, quantidade encomendada e localidade do cliente.*

Na *quantidade encomendada* não são considerados os tamanhos encomendados, i.e., numa encomenda, por exemplo de 50 unidades de uma calça com o código 4001202, não temos em atenção o tamanho encomendado.

Se o cliente for nacional, a *localidade* refere-se ao concelho onde existe o estabelecimento principal mas se o não for, a localidade corresponde ao respectivo país.

A partir daqueles atributos, obteve-se alguns atributos derivados, nomeadamente, *tipo* (correspondente à definição de família apresentada) e *variantes* (obtido pela associação entre variante de calça e casaco do produto).

6.2.2 Fase 2 - Estudo ao Longo de Três Momentos

Das regras obtidas ao longo dos três momentos em que decorreu este estudo e tendo em atenção que se pretendia evidenciar o processo de consolidação do conhecimento através da utilização do ZACCAR, seleccionámos⁹⁵:

- regras semelhantes obtidas nos três momentos;
- uma regra ocorrida no segundo momento a que não correspondia nenhuma regra semelhante no primeiro momento;
- uma regra ocorrida no terceiro momento a que correspondia uma regra semelhante no primeiro momento mas não no segundo;
- uma regra ocorrida no terceiro momento a que não correspondiam regras semelhantes nos dois momentos anteriores.

Para este estudo foi utilizado o algoritmo C5.0 [Clementine, 1997], www.rulequest.com, www.cs.umbc.edu/471/hw/hw11/c5tutorial.htm, do *Clementine* tendo, para o efeito, sido implementada uma *stream* conveniente (Figura 6.4).

⁹⁵ Por razões de confidencialidade não apresentamos, à semelhança do que foi feito no Caso 1, um anexo com as regras obtidas com o *Clementine* neste estudo de caso.

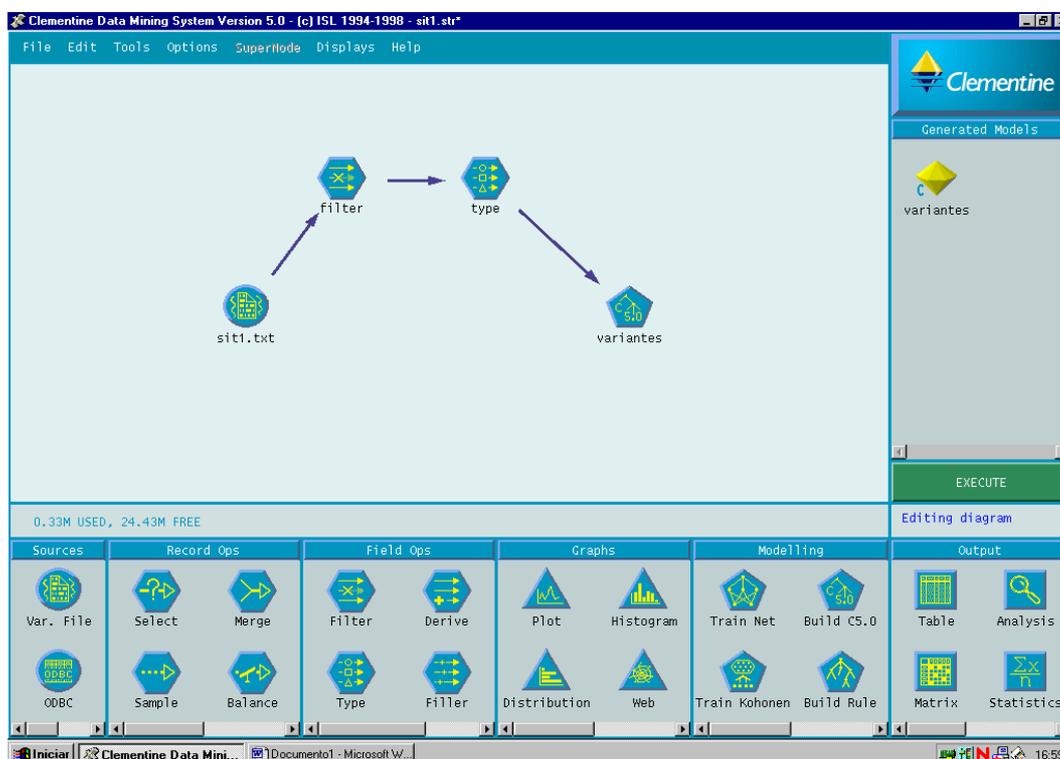


Figura 6.4 – Stream implementada no Clementine no segundo caso.

Momento 1

O momento 1 corresponde às encomendas anteriores às campanhas efectuadas, i.e., todos os dados registados entre o início da utilização da aplicação (1998) e 31 de Setembro de 2000. Neste período, foram encomendados 178418 produtos.

Os primeiros padrões deste estudo de caso, obtidos no momento 1, foram extraídos do ficheiro *padroes_dm* segundo a sintaxe do Clementine:

Rule #3 for calça com bainha :
if localidade == SÃO JOÃO DA MADEIRA
and tipo == calça
then -> calça com bainha (206, 0.769)⁹⁶

Rule #4 for calça com virola:
if localidade == LISBOA
and tipo == calça
then -> calça com virola (3586, 0.945)

⁹⁶ Estes valores representam, respectivamente, o número de ocorrências da base de dados que verificam os antecedentes da regra e o nível de confiança da regra num intervalo [0,1].

Rule #36 for calça com virola:
if localidade == BRAGA
and tipo == calça
then -> calça com virola (2820, 0,958)

Rule #1 for casaco c/1 abertura, calça com virola:
if localidade == PORTIMÃO
and tipo == fato com colete
then -> casaco c/1 abertura, calça com virola (205, 0.754)

Rule #1 for casaco c/2 aberturas, calça com virola:
if localidade == BRASIL
and tipo == fato
then -> casaco c/2 aberturas, calça com virola (3471, .,0)

Rule #60 for casaco c/2 aberturas, calça com virola:
if localidade == LISBOA
and tipo == fato
then -> casaco c/2 aberturas, calça com virola (5666, 0.872)

O procedimento Uniformização de Formatos, através da aplicação da função *unif*, transformou estas regras e colocou o resultado no ficheiro *temp1* conforme se mostra a seguir:

regra se localidade = "s. joão da madeira" e tipo = " calça"
entao
variantes= "calça com bainha"
parametros
[n1=1.18, n2= 76.9]

regra se localidade = "lisboa" e tipo = " calça"
entao
variantes= "calça com virola"
parametros
[n1=2.01, n2= 94.5]

regra se localidade = "braga" e tipo = " calça"
entao
variantes= "calça com virola"
parametros
[n1=7.58, n2= 95.8]

regra se localidade = "portimão" e tipo = " fato com colete"
entao
variantes= " casaco c/1 abertura, calça com virola"
parametros
[n1=0.11, n2= 75.4]

regra se localidade = "brasil" e tipo = " fato"
entao

variantes= " casaco c/2 aberturas, calça com virola"

parametros

[n1=1.95, n2= 100.0]

regra se *localidade = "lisboa" e tipo = "fato"*

entao

variantes= " casaco c/2 aberturas, calça com virola"

parametros

[n1=3.18, n2= 87.2]

Na fase de confronto do conhecimento e atendendo a que, neste primeiro momento, a BCCC não contém regras, a pesquisa efectuada não tem sucesso, havendo apenas uma transformação das regras canónicas em regras canónicas estendidas que são registadas no ficheiro *temp₂* (função *conf₂*).

Na etapa de participação do gestor do conhecimento, o conteúdo de *temp₂* é apresentado, regra a regra, ao gestor do conhecimento e este, optando por conservar todas as regras, segue sempre a opção 5 (Figura 5.7), i.e., introduz um comentário sobre a campanha levada a cabo. Cada regra, após esta intervenção, é registada em *temp₃*, tomando este ficheiro a forma:

regra se *localidade = "s. joão da madeira" e tipo = "calça"*

entao

variantes= "calça com bainha"

evolucao

[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.18,n2= 76.9))].

regra se *localidade = "lisboa" e tipo = "calça"*

entao

variantes= "calça com virola"

evolucao

[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=2.01,n2= 94.5))].

regra se *localidade = "braga" e tipo = "calça"*

entao

variantes= "calça com virola"

evolucao

[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=7.58,n2= 95.8))].

regra se *localidade = "portimão" e tipo = "fato com colete"*

entao

variantes= " casaco c/1 abertura, calça com virola"

evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=0.11,n2= 75.4))].

regra se localidade = "brasil" e tipo = "fato"

entao

variantes= " casaco c/2 aberturas, calça com virola"

evolucao

[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.95,n2= 100.0))].

regra se localidade = "lisboa" e tipo = "fato"

entao

variantes= " casaco c/2 aberturas, calça com virola"

evolucao

[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=3.18,n2= 87.2))].

Na assimilação do conhecimento, o conteúdo do ficheiro *temp₃* é transferido para a BCCC.

Momento 2

No momento 2 foram analisados os dados correspondentes às encomendas efectuadas de 1 de Outubro de 2000 a 31 de Janeiro de 2001, no total de 16809 registos; estes dados correspondem a um período em que estava a decorrer uma campanha que se traduziu na apresentação de faixas publicitárias durante os jogos de futebol da 1ª divisão.

Neste momento, obteve-se padrões semelhantes aos do momento anterior e alguns diferentes. Do conteúdo do ficheiro *padroes_dm* seleccionámos duas regras semelhantes a regras encontradas no momento anterior e um regra diferente:

Rule #1 for calça com virola:
if localidade == LISBOA
and tipo == calça
then -> calça com virola (266, 0.985)

Rule #2 for casaco c/2 aberturas, calça com virola:
if localidade == LISBOA
and tipo == fato
then -> casaco c/2 aberturas, calça com virola (371, 0.885)

Rule #16 for casaco c/2 aberturas, calça com virola:
if localidade == LISBOA

then -> casaco c/2 aberturas, calça com virola (1083, 0.304)

O procedimento Uniformização de Formatos transformou estas regras que foram registadas em *temp₁* conforme se mostra a seguir:

regra se *localidade = "lisboa" e tipo = "calça"*

entao

variantes="calça com virola"

parametros

[n1=1.58, n2= 98.5]

regra se *localidade = "lisboa" e tipo = "fato"*

entao

variantes=" casaco c/2 aberturas, calça com virola"

parametros

[n1=2.20, n2= 88.5]

regra se *localidade = "lisboa"*

entao

variantes=" casaco c/2 aberturas, calça com virola"

parametros

[n1=6.44, n2= 30.4]

Como as duas primeiras regras são semelhantes a outras já existentes, a pesquisa efectuada só não tem sucesso na última regra apresentada e, deste confronto, resulta um ficheiro *temp₂* cujo conteúdo é:

regra se *localidade = "lisboa" e tipo = "calça"*

entao

variantes= "calça com virola"

evolucao

*[(data(01,02,2001), comentario' ',parâmetros(n1=1.58,n2= 98.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das
campanhas'),parâmetros(n1=2.01,n2= 94.5))].*

regra se *localidade = "lisboa" e tipo = "fato"*

entao

variantes= " casaco c/2 aberturas, calça com virola"

evolucao

*[(data(01,02,2001), comentario' ',parâmetros(n1=2.20,n2= 88.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das
campanhas'),parâmetros(n1=3.18,n2= 87.2))].*

regra se *localidade = "lisboa"*

entao

variantes= " casaco c/2 aberturas, calça com virola"

evolucao

[(data(01,02,2001), comentario' ',parâmetros(n1=6.44,n2= 30.4))].

Pela análise do conteúdo de *temp*₂ pode-se verificar que:

- a segunda regra está contida na terceira;
- a terceira regra não foi detectada no momento 1.

E aqui surge uma primeira questão: atendendo à segunda regra, haverá algum interesse em reter a terceira?

Na etapa de participação do gestor do conhecimento, o conteúdo de *temp*₂ é apresentado, regra a regra, ao gestor do conhecimento e este introduz um comentário sobre a campanha que antecedeu este momento. Cada regra, após esta intervenção, é registada em *temp*₃, tomando este ficheiro a forma:

```
regra se localidade = "lisboa" e tipo = "calça"  
entao  
variantes= "calça com virola"  
evolucao  
[(data(01,02,2001), comentario('Análise após o início da 1ª campanha')  
,parametros(n1=1.58,n2= 98.5)),  
(data(01,10,2000), comentario('Análise das encomendas antes das  
campanhas'),parametros(n1=2.01,n2= 94.5))].
```

```
regra se localidade = "lisboa" e tipo = "fato"  
entao  
variantes= " casaco c/2 aberturas, calça com virola"  
evolucao  
[(data(01,02,2001), comentario('Análise após o início da 1ª campanha'),parametros(n1=2.20,n2=  
88.5)),  
(data(01,10,2000), comentario('Análise das encomendas antes das  
campanhas'),parametros(n1=3.18,n2= 87.2))].
```

```
regra se localidade = "lisboa"  
entao  
variantes= " casaco c/2 aberturas, calça com virola"  
evolucao  
[(data(01,02,2001), comentario('Análise após o início da 1ª campanha')  
,parametros(n1=6.44,n2= 30.4))].
```

Na fase de assimilação do conhecimento, a pesquisa não tem sempre sucesso pois algumas regras da BCCC não foram actualizadas; por esse motivo, pela actuação da função *pesquisa*, é criado o ficheiro *temp*₄ cujo conteúdo passa a ser:

```
regra se localidade = "s. joão da madeira" e tipo = "calça"  
entao
```

variantes= "calça com bainha"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.18,n2= 76.9))].

regra se *localidade = "braga" e tipo = "calça"*
entao
variantes= "calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=7.58,n2= 95.8))].

regra se *localidade = "portimão" e tipo = "fato com colete"*
entao
variantes= "casaco c/1 abertura, calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=0.11,n2= 75.4))].

regra se *localidade = "brasil" e tipo = "fato"*
entao
variantes= "casaco c/2 aberturas, calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.95,n2= 100.0))].

Em seguida, aquele procedimento transfere o conteúdo dos ficheiros *temp₃* e *temp₄* para a BCCC.

Momento 3

No momento 3, analisou-se os dados relativos ao período compreendido entre 1 de Fevereiro e 31 de Março de 2001, no total de 28594 registos. Neste intervalo de tempo, além da campanha iniciada no momento anterior, decorria uma outra campanha publicitária que patrocinou um dos programas de grande audiência da SIC (Sociedade Independente de Comunicação).

Neste momento, obteve-se padrões semelhantes aos dos momentos anteriores e alguns diferentes.

Do conteúdo do ficheiro *padroes_dm* seleccionámos, respectivamente, uma regra que apareceu pela primeira vez neste momento, uma regra que tinha sido detectada nos dois momentos anteriores e uma regra detectada no momento 1 mas não no momento 2:

Rule #4 for calça com bainha:

```
if localidade == SEIXAL
then -> calça com bainha (117, 0.387)
```

```
Rule #1 for calça com virola:
if localidade == LISBOA
and tipo == calça
then -> calça com virola (698, 0.97)
```

```
Rule #39 for calça com virola:
if localidade == BRAGA
and tipo == calça
then -> calça com virola (827, 0.353)
```

Estas regras, depois de transformadas em regras canónicas, foram registadas em $temp_1$ conforme se mostra a seguir:

```
regra se localidade = "seixal"
então
variantes=" calça com bainha"
parametros
[n1=0.41, n2= 38.7]
```

```
regra se localidade = "lisboa" e tipo = " calça"
então
variantes="calça com virola"
parametros
[n1=2.44, n2= 97.0]
```

```
regra se localidade = "braga" e tipo=" calça"
então
variantes=" calça com virola"
parametros
[n1=2.89 n2=,35.3]
```

Do confronto entre $temp_1$ e a BCCC, a pesquisa efectuada não tem sucesso na primeira regra, resultando, conseqüentemente, um ficheiro $temp_2$ cujo conteúdo é:

```
regra se localidade = "seixal"
então
variantes= "calça com bainha"
evolucao
[(data(01,06,2001), comentario' ',parâmetros(n1=0.41,n2= 38.7))].
```

```
regra se localidade = "lisboa" e tipo = " calça"
então
variantes= " calça com virola"
evolucao
[(data(01,06,2001), comentario' ',parâmetros(n1=2.44,n2= 97.0)),
```

*(data(01,02,2001), comentario('Análise após o início da 1ª campanha'),parâmetros(n1=1.58,n2=98.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=2.01,n2= 94.5))].*

regra se *localidade = "braga" e tipo="calça"*
entao
variantes= "calça com bainha"
evolucao
*[(data(01,06,2001), comentario(' ' ,parâmetros(n1=2.89,n2= 35.3)),
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=7.58,n2= 95.8))].*

Pela análise de *temp*₂, pode-se verificar que:

- a primeira regra só foi detectada neste momento;
- a segunda regra verificou-se nos três momentos, havendo ligeiras variações nos graus de suporte, tendo o grau de confiança um acréscimo com algum significado entre o primeiro (n2=94.5) e o segundo momento (n2=98.5), voltando a descer, embora menos, no terceiro momento (n2=97);
- a terceira regra foi detectada apenas nos momentos 1 e 3, estando a verificar-se uma diminuição no valor dos respectivos parâmetros (n1=7.58 e n2=95.8 no primeiro momento; n1=2.89 e n2=35.3 no segundo momento).

Na etapa de participação do gestor do conhecimento, o conteúdo de *temp*₂ é apresentado, regra a regra, ao gestor do conhecimento e este introduz um comentário sobre a campanha que antecedeu este momento. Cada regra, após esta intervenção, é registada em *temp*₃, tomando este ficheiro a forma:

regra se *localidade = "seixal"*
entao
variantes= "calça com bainha"
evolucao
*[(data(01,06,2001), comentario('Análise no decurso das duas campanhas')
,parâmetros(n1=0.41,n2= 38.7))].*

regra se *localidade = "lisboa" e tipo = " calça"*
entao
variantes= "calça com virola"
evolucao
[(data(01,06,2001), comentario('Análise no decurso das duas campanhas'),parâmetros(n1=2.44,n2= 97.0)),

*(data(01,02,2001), comentario('Análise após o início da 1ª campanha'),parâmetros(n1=1.58,n2=98.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=2.01,n2= 94.5))].*

regra se *localidade = "braga" e tipo="calça"*
entao
variantes= "calça com bainha"
evolucao
*[(data(01,06,2001), comentario('Análise no decurso das duas campanhas'),parâmetros(n1=2.89,n2= 35.3)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=7.58,n2= 95.8))].*

Na fase de assimilação do conhecimento, a pesquisa não tem sempre sucesso pois nem todas as regras da BCCC foram actualizadas; assim, o ficheiro *temp₄* é criado com as regras não actualizadas e o seu conteúdo passa a ser:

regra se *localidade = "lisboa" e tipo = "fato"*
entao
variantes= " casaco c/2 aberturas, calça com virola"
evolucao
*[(data(01,02,2001), comentario('Análise após o início da 1ª campanha'),parâmetros(n1=2.20,n2=88.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=3.18,n2= 87.2))].*

regra se *localidade = "lisboa"*
entao
variantes= " casaco c/2 aberturas, calça com virola"
evolucao
*[(data(01,02,2001), comentario('Análise após o início da 1ª campanha')
,parâmetros(n1=6.44,n2= 30.4))].*

regra se *localidade = "s. joão da madeira" e tipo = " calça"*
entao
variantes= "calça com bainha"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.18,n2= 76.9))].

regra se *localidade = "portimão" e tipo = " fato com colete"*
entao
variantes= " casaco c/1 abertura, calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=0.11,n2= 75.4))].

regra se *localidade = "brasil" e tipo = " fato"*
entao
variantes= " casaco c/2 aberturas, calça com virola"

evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.95,n2= 100.0))].

Pela assimilação dos ficheiros *temp*₃ e *temp*₄, o conteúdo da BCCC, no fim destes três momentos, passa a ser:

regra se *localidade = "seixal"*
entao
variantes= "calça com bainha"
evolucao
[(data(01,06,2001), comentario('Análise no decurso das duas campanhas') ,parâmetros(n1=0.41,n2= 38.7))].

regra se *localidade = "lisboa" e tipo = "calça"*
entao
variantes= "calça com virola"
evolucao
[(data(01,06,2001), comentario('Análise no decurso das duas campanhas'),parâmetros(n1=2.44,n2= 97.0)),
(data(01,02,2001), comentario('Análise após o início da 1ª campanha') ,parâmetros(n1=1.58,n2= 98.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=2.01,n2= 94.5))].

regra se *localidade = "braga" e tipo="calça"*
entao
variantes= "calça com bainha"
evolucao
[(data(01,06,2001), comentario('Análise no decurso das duas campanhas'),parâmetros(n1=2.89,n2= 35.3)),
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=7.58,n2= 95.8))].

regra se *localidade = "lisboa" e tipo = "fato"*
entao
variantes= " casaco c/2 aberturas, calça com virola"
evolucao
[(data(01,02,2001), comentario('Análise após o início da 1ª campanha'),parâmetros(n1=2.20,n2= 88.5)),
(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=3.18,n2= 87.2))].

regra se *localidade = "lisboa"*
entao
variantes= " casaco c/2 aberturas, calça com virola"
evolucao
[(data(01,02,2001), comentario('Análise após o início da 1ª campanha') ,parâmetros(n1=6.44,n2= 30.4))].

regra se *localidade = "s. joão da madeira" e tipo = "calça"*
entao

variantes= "calça com bainha"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.18,n2= 76.9))].

regra se *localidade = "portimão" e tipo = "fato com colete"*
entao
variantes= "casaco c/1 abertura, calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=0.11,n2= 75.4))].

regra se *localidade = "brasil" e tipo = "fato"*
entao
variantes= "casaco c/2 aberturas, calça com virola"
evolucao
[(data(01,10,2000), comentario('Análise das encomendas antes das campanhas'),parâmetros(n1=1.95,n2= 100.0))].

6.2.3 Discussão

Como foi possível observar, nos diferentes momentos do estudo, o novo conhecimento, na sequência do processo de consolidação de conhecimento levado a cabo, permite a actualização das regras existentes ou a sua eliminação, levando, assim, a que o conhecimento registado na BCCC se mantenha sempre actualizado face às rápidas modificações dos padrões de comportamento dos clientes.

O facto de se terem detectado, tal como no caso anterior, situações em que determinadas regras se podiam considerar contidas noutras mais gerais, leva-nos a concluir da necessidade de refinar o conceito de regras semelhantes apresentado na secção 5.1.1. de modo a que, no procedimento de assimilação, por um processo automático, fosse tomada a decisão de registar ambas as regra ou apenas uma delas (A mais particular? A mais geral?).

6.3 Situações não Previstas Detectadas no Estudo de Casos

O estudo destes dois casos permitiu a detecção de situações não previstas as quais mostram a necessidade de introdução de alguns melhoramentos no protótipo desenvolvido, nomeadamente:

- detecção de novas regras e o abandono de outras no lançamento de novas campanhas;
- a constatação de que o conhecimento descoberto pode ser redundante, levando à necessidade de refinar o conceito de regras de semelhança utilizado;
- situações relacionadas com aspectos temporais das regras pois certas regras que se tornaram obsoletas devem ser eliminadas da BCCC;
- tratamento de conflitos entre regras.

No processo desenvolvido, somos confrontados com três tipos distintos de regras em cada momento:

- regras que aparecem e que são confrontadas com outras já existentes, actualizando-as de acordo com o conceito de semelhança de regras introduzido;
- novas regras;
- regras existentes na BCCC que não são actualizadas.

As regras dos tipos 1 e 2 são, após a etapa de confronto, apresentadas ao gestor do conhecimento que tem possibilidade de as completar com comentários considerados pertinentes ou de as eliminar se considerar que já não interessam ao negócio em causa nem actual nem futuramente. Em qualquer uma das situações o gestor está-se a basear no seu conhecimento sobre o negócio e o seu domínio de aplicação.

As razões que o levam a eliminar as regras podem ser diversas como, por exemplo, constatar que uma regra, após um número n de observações, continua a manter os parâmetros de suporte e/ou confiança abaixo dum determinado nível.

No entanto e porque o gestor visualiza as regras uma a uma, não as podendo comparar, algumas regras que passaram nesta fase não deverão, eventualmente, ser registadas na BCCC por corresponderem a regras referentes a conhecimento redundante. Por essa razão, é oportuno o desenvolvimento de uma base de conhecimento do domínio que informe o assimilador como assimilar certas regras e eliminar outras, nomeadamente permitindo tratar questões ligadas a incoerências ou conflitos com as regras já existentes.

Esta base de conhecimento deve ser composta por um conjunto de restrições que, por exemplo, indiquem que de duas regras com o mesmo conseqüente e com os antecedentes duma sendo um subconjunto dos da outra, a primeira regra deve ser a única a ser registada.

Algumas das regras da BCCC que, no momento presente não foram confrontadas (regras do tipo 3), por uma questão de validade temporal devem, eventualmente, ser eliminadas por se terem tornado obsoletas.

Então, nas restrições definidas na base de conhecimento, esta situação deve, também, ser contemplada, tendo em atenção o âmbito do negócio (nalgumas situações, as regras na situação acabada de expor devem ser conservadas num ficheiro de histórico). Algumas situações possíveis para as regras são:

- não ter sido aumentado o seu campo de evolução na utilização do ZACCAR em n momentos consecutivos;
- referirem-se a um conjunto de produtos que deixaram de ser comercializados ou serviços já não existentes.

Para as regras deste tipo, há, ainda, outras situações como por exemplo:

- Novas regras serem semelhantes a estas, de acordo com a quantificação duma medida de semelhança cujo valor dependerá do número de antecedentes iguais nas duas regras. De acordo com a introdução desta medida, duas regras que sejam semelhantes conforme a definição introduzida no capítulo 5, dizem-se 100% semelhantes.

Exemplo:

regra se profissao_pai="10"

entao

opcao="S"

evolucao

*[(data (30,11,2000), comentario(`Depois da campanha de marketing 1 – Varadero+Havana`),
parâmetros (n1=3.922, n2=71))].*

e

regra se curso="1" e profissao_pai="10"

entao

opcao="S"

evolucao

*[(data (30,11,2000), comentario(`Depois da campanha de marketing 1 – Varadero+Havana`),
parâmetros (n1=2.241, n2=88))].*

Assim, os requisitos para o metaconhecimento a registar na base de conhecimento do domínio são:

- validade temporal das regras
- restrições:
 - ▶ regras com valores dos parâmetros de suporte e confiança inferiores a um determinado valor durante n momentos consecutivos;
 - ▶ quantificação duma medida de semelhança.

Pelo acabado de expor, consideramos a necessidade de quer o conteúdo do ficheiro $temp_3$ quer o de $temp_4$ deverem ser sujeitos às restrições definidas na base de conhecimento do domínio pelo que o procedimento assimilação descrito no capítulo 5 (Figura 5.8), deve ser alterado de acordo com o esquema da Figura 6.5.

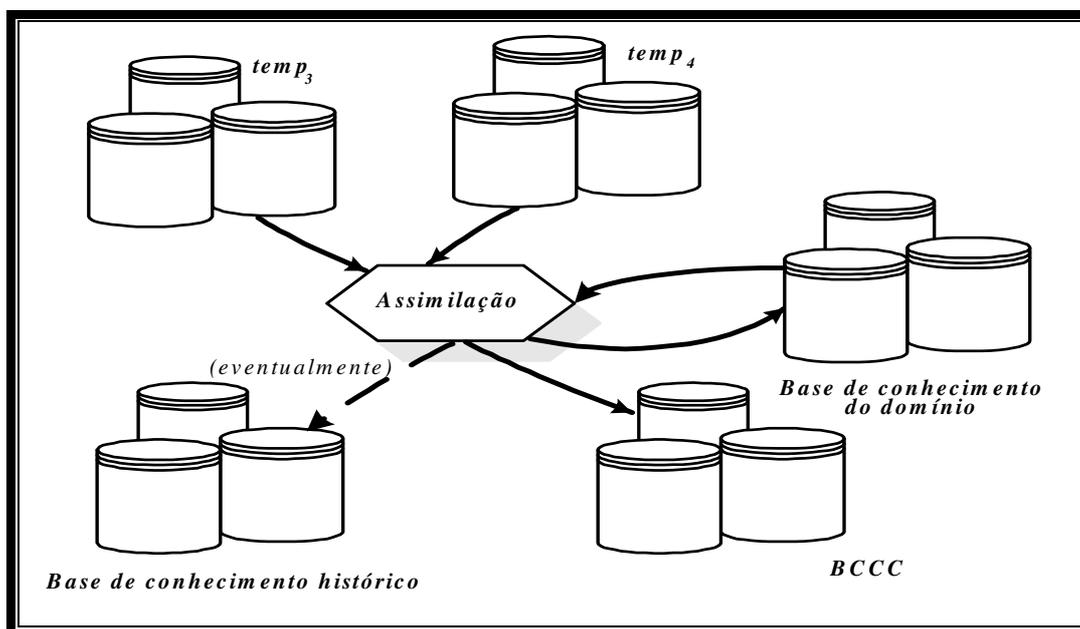


Figura 6.5 – Alteração a introduzir no procedimento "Assimilação".

6.4 Sumário

O protótipo desenvolvido foi aplicado em dois casos, tendo em vista explorar a viabilidade do sistema, analisando, ainda, pontos fracos e vantagens do desenvolvimento de novas funcionalidades.

No primeiro estudo, o ZACCAR foi aplicado numa experimentação simulada, sendo a principal preocupação, a partir da evolução das regras obtidas, a análise do impacto das campanhas de *marketing* consideradas.

No segundo estudo, que foi efectuado sobre os dados contidos nas bases de dados da AMMA, uma fábrica de confecção para homem, foi dada particular importância à análise do processo de consolidação do conhecimento descoberto, uma funcionalidade chave do ZACCAR.

Ambos os estudos mostraram que o ZACCAR tem boas potencialidades que se poderão tornar muito úteis em qualquer empresa onde seja feita a sua implantação quer

como sistema independente quer como integrado noutros sistemas informáticos de maior abrangência.

7. Conclusões

“To reflect is to look back over what has been done so as to extract the net meanings that are the capital stock for intelligent dealing with further experiences. It's the heart of intellectual organization and of the disciplined mind ” (John Dewey⁹⁷)

Este capítulo enquadra e caracteriza, de uma forma resumida, o trabalho de pesquisa e desenvolvimento efectuado, referenciando os atributos mais significativos e apontando caminhos para trabalho futuro.

O trabalho presente recebe inspiração de vários campos do saber como as Ciências da Computação, a Matemática, a Lógica, a Programação em Lógica, a Inteligência Artificial e aplica um conjunto de técnicas, nomeadamente as bases de dados, os *data warehouses*, a DCBD, o DM e o CRM.

7.1 Síntese do Trabalho Desenvolvido

Após uma observação da situação em que os negócios se desenrolam, concluímos que a gestão do relacionamento com os clientes é uma das principais práticas de gestão

⁹⁷ John Dewey (1859 - 1952) é, provavelmente, quem deu o contributo mais significativo para o desenvolvimento do pensamento educacional no século XX - apud www.infed.org/thinkers/et-dewey.html. Consultado em 17/01/2001.

e que, para que tal seja possível, as empresas têm que se preocupar em zelar pelo Conhecimento do Comportamento dos Clientes.

Na sequência desta conclusão, começámos por fazer uma análise cuidadosa sobre as vantagens do CCC e de uma gestão do relacionamento com os clientes.

De acordo com leituras realizadas e observação de situações reais, constatámos a existência de ferramentas para recolha dos dados das compras dos clientes, solicitações de serviços ou outros contactos cliente-empresa e sua transformação em conhecimento conceptual. Este conhecimento existe sob uma forma estática, não acompanhando o ambiente dinâmico em que os clientes efectuem as suas compras o que altera, com muita frequência, os padrões de comportamento já conhecidos.

Deparámo-nos, assim, com uma lacuna no processo: o zelar pelo conhecimento adquirido tendo em vista uma melhor actuação da organização. Para que tal fosse possível, propusémo-nos desenvolver uma sistema de conhecimento - o sistema ZACCAR - responsável por manter uma base de conhecimento do comportamento dos clientes e formado, ainda, por um módulo que potenciasses a actualização dinâmica desse conhecimento através dum processo de aprendizagem semi-automático e incremental.

Uma função chave do ZACCAR é a atenção que é dada ao processo de consolidação do conhecimento descoberto. Trata-se de um processo que incorpora novo conhecimento no conhecimento organizacional, registando-o numa base de conhecimento através da confrontação com outro conhecimento previamente aí registado. Esta assimilação do conhecimento inclui informação sobre a evolução temporal e o tratamento dos conflitos que possam surgir.

A arquitectura do sistema incorpora:

- uma Base de Conhecimento do Comportamento dos Clientes
 - um Gestor de Conhecimento do Comportamento dos Clientes
- e por razões óbvias, um módulo de consulta à BCCC.

Um protótipo do sistema foi desenvolvido, usando o formalismo de representação do conhecimento correspondente a regras de produção estendidas e a linguagem de programação em lógica *Prolog* para implementação (Anexo 3).

A exequibilidade do sistema é aferida através de dois estudos de casos usando o protótipo desenvolvido.

O primeiro estudo desenvolveu-se a partir de uma experiência simulada, usando os dados coleccionados a partir de campanhas de *marketing* sobre uma viagem de fim de curso (Anexos 4 e 5).

Ao longo de quatro momentos, foi estudada a evolução de duas regras seleccionadas para o estudo (Anexo 6), permitindo realçar duas características importantes do sistema:

- a possibilidade de actualizar o CCC de uma forma automática;
- a facilidade de consulta de uma base de conhecimento dinamicamente actualizada onde os membros da organização podem observar a evolução dos padrões de comportamento, representados sob a forma de regras.

Para além destes aspectos, os padrões e a sua evolução provaram a sua utilidade para:

- a análise do impacto de campanhas de *marketing*;
- a orientação na gestão de novas actividades de CRM;
- a detecção de diferenças significativas entre os parâmetros de um mesmo padrão em diferentes momentos e a possibilidade de uma análise e registo da evolução dessas diferenças.

O segundo estudo utilizou os dados contidos nas bases de dados transaccionais da AMMA (Anexo 7), uma fábrica de confecção para homem.

Neste estudo foi dada particular importância à análise do processo de consolidação do conhecimento descoberto, uma funcionalidade chave do ZACCAR.

À semelhança do estudo anterior, também este se desenvolveu ao longo de vários momentos, coincidindo cada momento com campanhas de *marketing* efectuadas pela AMMA.

Como foi possível observar, a partir de algumas regras seleccionadas, o novo conhecimento permite a actualização das regras existentes ou a sua eliminação, levando a que o conhecimento registado na BCCC se mantenha sempre coerente e actualizado.

No fim deste estudo foi possível apontar uma capacidade importante do ZACCAR - a consolidação de novo conhecimento face a conhecimento já registado.

As conclusões obtidas com os estudos de casos, utilizando o protótipo desenvolvido, são encorajadoras e sugerem que o sistema tem boas potencialidades que, se convenientemente exploradas, poderão ser muito úteis.

É nossa convicção que a introdução daquele sistema numa empresa permitirá a disponibilização de conhecimento actualizado, completo e retratando a evolução temporal dos padrões de comportamento. Este conhecimento contribuirá para a realização de campanhas de *marketing* baseadas em informação real e actualizada além de outras acções que tenham em vista alterar os padrões de compra dos clientes ou melhorar as condições oferecidas pela organização aos seus clientes.

O facto deste sistema ter sido desenvolvido em módulos, não sendo, por isso, encarado como um sistema fechado, abre, ainda, a possibilidade da sua integração noutros sistemas informáticos de maior abrangência.

O trabalho desenvolvido foi já objecto de publicações parcelares que pretendem dar um contributo para o conhecimento científico nas áreas abrangidas por este trabalho e que foram:

Gonçalves, M. F.; Carvalho, J. A. and Santos, M. F. – “A Customers’ Behaviour Knowledge Repository”, in Hackney, R. (Ed.), *Proceedings of the Bit’99 Conference*, Faculty of Management and Business, The Manchester Metropolitan University, Manchester: 1999 (edição em CD-ROM).

Gonçalves, M. F.; Carvalho, J. A.; Santos, M. F. - “Gestão do Conhecimento do Comportamento dos Clientes: Utilização de *Data Mining* e Base de Conhecimento”, *Actas das X Jornadas Luso-Espanholas de Gestão Científica*, Volume V- Gestão Comercial e *Marketing*, Universidade do Algarve, Vila Moura: Fevereiro de 2000, pp. 81-7.

Gonçalves, M. F.; Carvalho, J. A.; Santos, M. F. - “Um sistema de apoio à gestão do conhecimento do comportamento dos clientes”, in Amaral, L. e Carvalho, J.A. (Ed.), *Actas da 1ª Conferência da Associação Portuguesa dos Sistemas de Informação*, Guimarães: 25-27 de Outubro 2000, Artigo S4B2 (edição em CD-ROM).

Gonçalves, M. F.; Carvalho, J. A.; Santos, M. F. - “A Knowledge System for *CRM*”, a aparecer em *Proceedings of BITWorld 2001 Conference*, Cairo: Junho 2001.

Gonçalves, M. F.; Carvalho, J. A.; Santos, M. F. - “CRM - Processo de Consolidação do Conhecimento Descoberto em Bases de Dados”, submetido para apresentação na *2ª Conferência da Associação Portuguesa dos Sistemas de Informação*, a ter lugar em Évora em Outubro de 2001.

7.2 Perspectivas futuras

No âmbito do trabalho ora desenvolvido, com um horizonte temporal bem delimitado, não foi possível aperfeiçoar alguns dos procedimentos desenvolvidos ou introduzir novas capacidades que reputamos de interessantes e úteis.

A atenção esteve centrada na viabilidade do processo de consolidação do novo conhecimento e sua integração noutro conhecimento já existente, não tendo sido dado grande ênfase a questões de aperfeiçoamento técnico ou a funcionalidades mais dirigidas para aspectos particulares.

O protótipo foi desenvolvido única e exclusivamente com a preocupação de demonstrar a viabilidade e eficácia do sistema de conhecimento proposto.

A utilização do ZACCAR em ambientes reais, permitirá, concerteza, que melhor se apure o seu desempenho e potencialidades efectivas.

É, contudo, possível desde já antever que o sistema é passível dum conjunto de melhoramentos a ter em atenção:

- O procedimento "*Interface*" (Figura 5.6), integrado na etapa *Interface* do GCCC, poderá ser melhorado no sentido de permitir ao gestor do conhecimento a possibilidade de, face à análise de uma regra, realizar tarefas complementares, nomeadamente, a análise directa dos dados contidos nos *data warehouses*, através da execução de comandos SQL ou outras tarefas afins.
- O desenvolvimento de novas facilidades de pesquisa no módulo "Consulta à BCCC", permitindo, por exemplo, a pesquisa a partir dum conjunto de antecedentes.
- Os dois *interfaces* que integram o ZACCAR (Figura 5.6 e Figura 5.10) foram desenvolvidos apenas de modo a contribuir para a demonstração da exequibilidade do sistema e sem quaisquer preocupações de usabilidade. Deste modo, em futuros desenvolvimentos do sistema, é necessária a sua melhoria numa perspectiva das boas práticas de *Human Computer Interface*⁹⁸ (HCI) [Marciniak, 1994], [Gulliksen & Sadblad, 1995], [Gulliksen, 1996], [Toleman, 1996].
- Tornar mais lato o conceito de regras semelhantes apresentado em 5.1.1, levando à quantificação de uma medida de semelhança através da implementação de uma base de conhecimento do domínio que informe o assimilador como deverá assimilar certas regras e eliminar outras, usando operações de generalização ou particularização das regras. Esta necessidade foi detectada por se terem obtido regras que se podiam considerar englobadas noutras mais gerais. Pretende-se, também, nesta base de conhecimento, contemplar situações relacionadas com aspectos temporais, como critérios de durabilidade das regras o que permitirá que o sistema se encarregue, por exemplo, da eliminação de regras da BCCC obsoletas que não foram passíveis de confronto com novas regras (Figura 6.5).

⁹⁸ *HCI* é uma tecnologia relacionada com o desenho, desenvolvimento e implementação de sistemas computacionais interactivos para utilização humana, tendo em atenção o estudo de outros fenómenos que os envolvem [Hewett *et al.*, 1997].

- O desenvolvimento de novas funcionalidades necessárias a um melhor desempenho do sistema, nomeadamente:
 - ▶ um sistema de inferência que, actuando a partir das regras registadas na BCCC, permita novos relacionamentos;
 - ▶ o apoio interactivo a campanhas de *marketing* e a análise da influência dessas campanhas no comportamento dos clientes incorporando, de qualquer modo, o resultado dessa análise na base de conhecimento;
 - ▶ a instalação dum sistema de recomendações [Kudyba & Hoptroff, 2001], [Schafer *et al.*, 2001], principalmente orientado para o comércio electrónico que, em ligação directa com o *browser* a utilizar pelo cliente e a BCCC, face à pesquisa dum determinado produto pelo cliente, sugira outros produtos com aquele relacionados.

Provavelmente, em determinadas empresas, serão consideradas vantajosas várias ocorrências do ZACCAR, conduzindo à criação e manutenção de diferentes BCCC para cada produto ou área do negócio havendo, nesse caso, a necessidade de um sistema a nível superior que faça a agregação dos diversos sistemas.

Em tais condições, a fase do confronto tem que ser modificada de modo a possibilitar que, para cada nova regra, seja feita uma selecção da base de conhecimento a pesquisar para o procedimento de confronto e posterior assimilação do novo conhecimento.

Referências Bibliográficas

- Agrawal, Rakesh; Imielinski, Tomasz and Swami, Arun - "Database Mining: a Performance Perspective", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 5, nº 6, December 1993, pp 914-925.
- Aho, Alfred V.; Sethi, Ravi; Ullman, J.D. - "*Compilers: Principles, Techniques and Tools*", Addison_Wesley Publishing Company, Menlo Park, California:1985.
- Anand, T.; Kahn, G. – “ Focusing Knowledge-Based Techniques on Market Analysis“, *IEEE Expert*, Vol. 8, ° 4 (August 1993), pp. 19-24.
- Baião, Rui – “E-Telcos”, *Unisys Newsletter*, Especial Novembro-Dezembro 2000.
- Berry, Michael J. A.; Linoff, Gordon S. – “*Data Mining Techniques – For Marketing, Sales and Customer Support*”, John Wiley & Sons, Inc., New York:1997.
- Berry, Michael J. A.; Linoff, Gordon S. – “*Mastering Data Mining*”, John Wiley & Sons, Inc., USA:2000.
- Berson, Alex; Smithy Stephen; Thearling, Kurt - "*Building Data Mining Applications for CRM*", McGraw-Hill, New York: 1999.
- Bits & Pixels – “*Customer Intelligence Agency – Understanding* “. Disponível na Internet: <URL: www.bitpix.com/cia/fraud.htm>. Consultado em 25/01/2001.
- Blackwell, Roger D., Miniard, Paul W.; Engel, James F. – “*Consumer Behavior*“, Ninth edition, Harcourt College Publishers, New York: 2001.
- Brachman, Ronald J.; Anand, Tej – “*The Process of Knowledge Discovery in Databases*” in [Fayyad *et al.*, 1996], pp 37-57
- Brachman, Ronald J.; Khabaza, Tom; Klöesgen, Willi; Pietetsky-Shapiro, Gregory and Simoudis, Evangelos – “Mining Business Databases”, *Communications of the ACM*, Vol. 39, nº 11 (Nov. 1996), p 43-8.

- Brackett, Michael H. – “Business Intelligence Value Chain”, *DM Review*: March 1999.
- Brathwaite, K.S. – “Information Engineering: Development Issues”, CRC Press, 1992.
- Bratko, I. – “*Prolog: Programming for Artificial Intelligence*”, (2nd ed.), Addison-Wesley, New York:1990.
- Brito, Carlos Melo – “Os novos desafios do *marketing* da terceira vaga”, *Actas das X Jornadas Luso-Espanholas de Gestão Científica*, Vol.5, Vilamoura: 2000, pp. 11-29.
- Brown, Stanley A. - “*Customer Relationship Management*”, John Wiley & Sons Canada, Lda.Ontario:2000.
- Burnett, Ken – “*The handbook of key customer relationship management*”, Prentice Hall, London:2001.
- Cabena, P.; Hadjinian, P.; Stadler, R.; Verhees, J.; Zanasi, A. – “*Discovering Data Mining From Concept to Implementation*”, Prentice Hall PTR, Upper Saddle River, New Jersey 07458:1997.
- Cabibbo, Luca; Torlone, Riccardo - "*Querying Multidimensional Databases*", Dipartimento di Informatica e Automazione, Università di Roma Tre, Disponível na Internet: <URL: www.dia.uniroma3.it/~cabibbo/pub/pdf/dbp197.pdf>. Consultado em 19-03-2001.
- Cardoso, Domingos Moreira - "*Programação Lógica e Demonstração Automática de Teoremas*", Cadernos de Matemática CM/D-03-Dep. de Matemática da Universidade de Aveiro, Aveiro:1995.
- Carvalho, J. A. ; Morais, P. - "Sistemas Informáticos e Conhecimento Organizacional: Uma Reinterpretação dos Papéis Desempenhados pelos Sistemas Informáticos nas Organizações". Aceite para apresentação na 2ª Conferência de Sistemas de Informação a realizar em Évora, em Dezembro de 2001.
- Cassel, Catherine and Symon, Gillian – “*Qualitative Methods in Organizational Research*”, SAGE Publications, Third Edition, London:1997.
- Cavaye, A. L. M. – “Case study research: a multi-faceted research approach for IS”, *Info Systems J.* (1996) 6, pp 227-42.
- Choo, C.W. – “The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions”, *Int. Journal of Information Management*, Vol. 16, Nº 5 (1996): 329-340. Disponível na Internet: <URL: choo.fis.utoronto.ca/FIS/ResPub/Koart.html>. Consultado em 20/02/1998.
- "Clementine, Data Mining Systems Version 4 – User Guide", Integral Solutions Limited. Basingstoke:1997.
- Clocksinn, W.F.; Mellish, C.S., - "*Programming in Prolog*", Springer-Verlag: 1987.
- Codd, E.F. – “A Relational Model for Large Shared Data Banks”, *Communications of the ACM*, Vol. 13, nº 6 (1970), pp 377-87.

- Coelho, H.; Cotta, J.C. - "*Prolog by Example, How to Learn, Teach and Use it*", Springer-Verlag, 1988.
- Curry, Jay; Curry, Adam – "*The Customer Marketing Method*", The Free Press, New York:2000.
- Date, C.J. – "*An Introduction to Database Systems*", Addison-Wesley Publishing Company, 6th edition, Merlo Park, California:1995.
- Demarest, Marc – "Building the Data Mart", *DBMS*, Volume 7, Number 8 (July 1994).
- Dhar, Vasant; Stein, Roger – "*Seven Methods for Transforming Corporate Data into Business Intelligence*", Prentice Hall, Upper Saddle River, New Jersey: 1997.
- Dougherty, Ray C. – "*Natural Language Computing*", Lawrence Erlbaum Associates, Inc. Publishers, Hillsdale, New Jersey: 1994.
- Etzioni, Amitai – "*Organizações Modernas*", Livraria Pioneira, 6ª Edição, S. Paulo:1980.
- Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy (Eds.) – "*Advances in Knowledge Discovery and Data Mining*". AAAI Press/The Mit Press, Menlo Park, California: 1996.
- Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic – "*From Data Mining to Knowledge Discovery: an Overview*" in [Fayyad *et al.*, 1996], pp 1-34.
- Fayyad, Usama M.; Haussler, David; Stoplorz, Paul – "Mining Scientific Data", *Communications of the ACM*, New York, Vol.3, nº 11, (November 1996), pp 51-7.
- Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic – "The KDD Process for Extracting Useful Knowledge from Volumes of Data", *Communications of the ACM*, New York, Vol.3, nº 11, (November 1996), pp 27-34.
- Fayyad, Usama M. – "Data Mining and Knowledge Discovery: Making Sense Out of Data", *IEEE Expert - Intelligent Systems & Their Applications*, Vol. 11, nº 5 (October 1996), pp 20-5.
- Feldens, Miguel – "*Integrando Data Mining em Call Centres*" – Edição de 072000 de Data Mining BR, em datamining.br.listbot.com
- Ferrão, Francisco – "*Webmining (o data mining da Internet)*" – Suplemento do Jornal Expresso de 27/01/2001.
- Figher, D. H. – "Knowledge Acquisition Via Incremental Conceptual Clustering", *Machine Learning*, 2, (1987), pp 139-72.
- Franks, R.V. – "*Computing Terminology*", 4ª Ed., CIMA, London:1997.

- Frawley, William J.; Piatetsky-Shapiro; Gregory; Matheus, Christopher, J. – “*Knowledge Discovery in Databases: an Overview*”, in [Piatetsky & Frawley., 1991], pp 1-27.
- Freitas, Alex A.; Lavington, Simon H. – “*Mining Very Large Databases with Parallel Processing*”, Keuwer Academic Publishers, London:1998.
- Frost, R.A. - “*Introduction to Knowledge Base Systems*”, William Collins Sons & Co. HD, 2ª Ed., London:1987.
- Furtado, Paula - “*Storage Management of Multidimensional Arrays in Database Management Systems*”, Shaker Verlag, Aachen: 2000.
- Garshol, Lars Marius – “*BNF and EBNF: What are they and how do they work?*”. Disponível na Internet: <URL: www.garshol.priv.no/download/text/bnf.html#id1.2>. Consultado em 31/01/2001.
- Groth, Robert – “*Data Mining – Building Competitive Advantage*”, Prentice Hall, Inc., New Jersey:2000.
- Gulliksen, Jan – “*Design for Usability - Domain-Specific Human-Computer Interfaces in Working Life*”. Acta Univ. Ups., Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, Uppsala:1996.
- Gulliksen, Jan; Sandblad, Bengt – “Domain Specific Design of User Interface”, *International Journal of Human-Computer Interaction*, Vol. 1, Nº 1, pp. 135-151:1995.
- Gutiérrez, S.S.M., Cillán, J.G.; Izquierdo, M.C.C. – “Una Propuesta del Processo de Decisión al en los Mercados de Consumo”, *Actas das X Jornadas LusoEspanholas de Gestão Científica*, Vol.V, pp 511-518, Vila Moura, Portugal: Fev.2000.
- Han, Jiawei; Kamber, Micheline - “*Data Mining - Concepts and Techniques*”, Academic Press, Morgan Kaufmann Publishers, San Francisco, U.S.A., 2001.
- Hand, David J. – “Statistical Expert Systems”, *Chance*, Vol.7-Nº 1, 1994, pp 28-34.
- Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strnog and Verplank – “ACM SIGCHI Curricula for Human-Computer Interaction”, 1996. Disponível na Internet: <URL: sigchi.org/cdg/index.html>. Consultado em 31/01/2001.
- Holsheimer, Marcel; Kersten, M. L. - “*Architectural Support for Data Mining*”, Computer Science/Department of Algorithmics and Architecture; Centrumm voor Wiskunde en Informatica, CS-R9429, Amsterdam:1994.
- Holsheimer, Marcel – “Strategic Data Mining”, Seminário “*Decision Support & Data Warehouse*”, Lisboa: 25-26 de Outubro de 1995.

- Hong, Jiarong; Mao, Chengjing - "*Incremental Discovery of Rules and Structure by Hierarchical and Parallel Clustering*", in [Piatetsky & Frawley, 1991], pp 177-94.
- Howe, Denis – "*The Free Online Dictionary of Computing*", 1995 . Disponível na Internet: <URL: burks.bton.ac.uk/burks/foldoc/36/40.htm>. Consultado em 31/01/2001.
- Hu, Xiaohua – "*Knowledge Discovery in Databases: an Attribute-Oriented Rough Set Approach*". Thesis for the Degree of Doctor of Philosophy in Computer Science, Faculty of Graduate Studies, University of Regina:1995.
- Inmon, W. H. – "*Building the Data Warehouse*", QED Technical Publishing Group, Wellesley:1992.
- Inmon, W. H.; Hackathorn, D. – "*Using the Data Warehouse*", John Wiley & Sons, Inc., New York:1994.
- Inmon, W. H. – "*The Data Warehouse and Data Mining*", *Communications of the ACM*, New York, Vol.3, nº 11, (November 1996), pp49-50.
- Kenan Technologies – "*An Introduction to Multidimensional Database Technology*" , Kenan Systems Corporation, 1995.
- Klösigen, Willi; Zytkow, Jan M. – "*Knowledge Discovery in Database Terminology*" in [Fayyad *et al.*, 1996], pp 573-92.
- Kohavi, R.; Provost, F. – "Applications of Data Mining to Electronic Commerce", Special Issue of the International Journal "*Data Mining and Knowledge Discovery*", February 2001. Disponível na Internet <URL: robotics.stanford.edu/~ronnyk/ecommerce-dm >. Consultado em 13/10/2000.
- Krogh, George Von; Nonaka, Ikujiro; Nischiguchi, Toshihino – "*Knowledge creation – A source of value*", St.Martin's Press, Inc.. New York:1999.
- Kudyba, Stephan; Hoptroff, Richards - "*Data Mining and Business Intelligence*", Idea Group Publishing, 2001.
- Lopes, F.; Morais, P.; Carvalho, J.A.; Amaral,L. – "A Systematization of Knowledge to support an Organisational Repository", in Hackney, R. (Ed.), *Proceedings of the Bit'99 Conference*, Faculty of Management and Business, The Manchester Metropolitan University, Manchester: 1999 (edição em CD-ROM).
- Lucas, Ana – "*Conferência de Sistemas de Suporte à Decisão & Data Warehouse*", Universidade Portucalense, Porto: Maio 1997.
- Lugger, G. F.; Stubblefield, W.A. - "*Artificial Intelligence Structures and Strategies for Complex Problem Solving*", Addison-Wesley, 1998.

- Mannila, Heikki; Toivonen, Hannu - "Discovering generalized episodes using minimal occurrences" , *Proceedings The Second International Conference on Knowledge Discovery & Data Mining*, Edited by Evangelos Simoudis, Jiawei Han & Usama Fayyad, AAAI Press Menlo Park, California: 1996, pp 146-51.
- Marciniak, John J. - "*Encyclopedia of Software Engineering*", John Wiley & Sons, Inc. New York, 1994.
- Mason, Tony; Brown, Duog - "*lex & yacc*", O'Reilly & Associates, Inc. USA, 2ª ed., 1991.
- Mello, F. Galvão – "*Probabilidades e Estatística*", Escolar Editora, Lisboa: 1997.
- Montcel, Henri Tezenas – "*Dicionário de Gestão*", Publicações D. Quixote, Lisboa: 1972.
- Moriarty, Terry – *Iemagazine*, October 5; 1999, Vol.2 ,nº 14.
- Newell, Frederick – "*loyalty.com – Customer Relationship Management into the New Era of Internet Marketing*", McGraw-Hill, New York: 2000.
- Newell, Frederick - "Ame os seus clientes", *executivedigest*, Ano 6, nº 71 (Setembro 2000), p 42-8.
- Nonaka, I., Takeuchi, H. – "*The Knowledge-Creation Company*", Oxford University Press, Oxford: 1995
- OLAP Council. OLAP Council White Paper: 1997. Disponível na Internet <URL: www.olapcouncil.org/research/whtpaply.htm>. Consultado em 10/10/2000.
- Okell, Janine Neural – "*Networks versus Chaid*", A White Paper from SmartFOCUS. Disponível na Internet <URL: www.crm-forum.com/crm_forum_white_papers/nnvc/sld01.htm>. Consultado em 4/9/2000.
- Parsaye, Kamran – "Surveying Decision Support: New Realms of Analysis", *Database Programming & Design on-line*: 1997. Disponível na Internet <URL: www.dbpd.com/parsaye.htm>. Consultado em 10/02/97.
- Parsaye, Kamran; Chignell, Mark - "*Intelligent Databases Tools & Applications*", John Wiley & Sons, Inc., New York: 1993.
- Peppers, Don; Rogers, Martha - "O triunfo da intimidade", *executivedigest*, Ano 6, nº 71 (Setembro 2000), pp 51-6.
- Pereira, J. L. – "*Tecnologia de Bases de Dados*", FCA Editora de Informática, 2ª Edição, Lisboa: 1998.
- Piatetsky-Shapiro, Gregory; Frawley, William J. (Eds) – "*Knowledge Discovery In Databases*", AAAI Press/The MIT Press, Menlo Park, California: 1991.
- Prusak, Laurence - "*Knowledge in Organization*", British Library Cataloging-in-Publication Data, USA: 1997

- Rodrigues, Fátima; Henriques, Pedro Rangel; Santos, Maribel - "Descoberta do Conhecimento em Bases de Dados - Tutorial", in Amaral, L. E Carvalho, J.^a (Ed.), *Actas da 1ª Conferência da Associação Portuguesa de Sistemas de Informação*, Guimarães, 2000 (edição em CD-ROM).
- Santos, M.F.V.T. - "*Sistemas de Classificação em Ambientes Distribuídos*", Tese de Doutoramento, Universidade do Minho, Escola de Engenharia, Departamento de Informática: 1999.
- Schafer, J. Ben; Konstan, Joseph A.; Riedl, John – “E-commerce Recommendation Applications” – Special issue of the International Journal “*Data Mining and Knowledge Discovery*“, February 2001. Disponível na Internet <URL: robotics.stanford.edu/~ronnyk/ecommerce-dm/>. Consultado em 13/10/2000.
- Scott, P. Nicolson, R. - "*Cognitive Science Projects in Prolog*", Lawrence Erlbaum Associates Publishers, USA: 1991.
- Shindler, Daniel - "Echocardiographic Data Mining", *E-chocardiography Journal, An Electronic Journal of Cardiac Ultrasound*, Set.5,1997. Disponível na Internet <URL: www2.umdj.edu/~shindler/datamine.html>. Consultado em 04/01/1998.
- Shinghal, R. - "*Formal Concepts in Artificial Intelligence*", Chapman & Hall, 1992.
- Shoam, Y - "*Artificial Intelligence Techniques in Prolog*", Morgan Kaufmann Publishers, USA: 1994.
- Silberschatz, A.; Stonebraker, M. and Ullman, J. D. - "Achievements and opportunities", *CACM*, Vol.34, nº 10, 1995, Report of an NSF Workshop on the Future of Database Systems Research, May 26/27-1995. Disponível na Internet <URL: dbpubs.stanford.edu:8090/pub/1995-15>. Consultado em 16/11/2000.
- Silberschatz, Abraham; Korth, H. F.; Sudarshan, S. - "*Database System Concepts*", Third Edition, WCB/McGraw-Hill:1996.
- Silberschatz, Avi; Tuzhilin, Alexander - "*User-Assisted Knowledge Discovery: How much should the User be Involved*", Position Paper, Bell Laboratories, 600 Montain Avenue, Murray Hill, NJ:1996.
- Simoudis, Evangelos; Kerber, Randy – “*Integrating Inductive and Deductive Reasoning for Data Mining*”, in [Fayyad *et al.*, 1996], pp 353-73.
- Simoudis, Evangelos - "Reality Check for Data Mining", *IEEE Expert*, Vol. 14, nº5, October 1996. Disponível na Internet <URL: church.computer.org/intelligent/ex1996/x5026abs.htm>. Consultado em 16/11/2000.

- Stefix, Mark – “*Introduction to Knowledge Systems*”, Morgan Kaufmann Publishers, Inc., San Francisco, California: 1995.
- Sterling, L. and Shapiro, E. - “*The Art of Prolog: Advanced Programming Techniques*”, Edited by Mit Press 2nd edition, 1994.
- Stonebraker, Michael; Rowe, Lawrence A.; Lindsay, Bruce G.; Gray, Jim; Carey, Michael J.; Brodie, Michael L.; Bernstein, Philip A.; Beech, David - “*Third -Generation Database System Manifesto - The Committee for Advanced DBMS Function*”. , Vol. 19, nº 3, 1990, pp 31-44.
- Thierauf, Robert J. – “*On-Line Analytical Processing Systems for Business*”, Quorum Books, Westport, Connecticut:1997.
- Toleman, Mark A. – “*The Design of the User Interface for Software Development Tools*”, PhD Thesis, Department of Computer Science at the University of Queensland, 1996.
- Ullman, Jeffrey D. – “*Principles of Database Systems*”, Rockville: Computer Science Press Corp, Second Edition, USA:1982.
- Uthurusamy, Ramasamy – “*From Data Mining to Knowledge Discovery; Current Challenges and Future Directions*”, in [Fayyad et al., 1996], pp 561-9.
- Vogel, Peter – “Know your Business; Build a Knowledgebase”, *Datamation*, July 1996.
- Wand, Yair; Monarchi, David E.; Parsons, Jeffrey; Woo, Carson C. – “Theoretical foundations for conceptual modelling in Information Systems development”, *Decision Support Systems*, 15:1995, pp 285-304.
- Wand, Yair; Wang, Ricahrd Y.– “Anchoring Data Quality Dimensions in Ontological Foundations”, *Communications of the ACM*, New York, Vol.3, nº 11, (November 1996), pp 86-95.
- Zorrinho, Carlos – “*Gestão da Informação*”, Editorial Presença Limitada, Lisboa:1991.

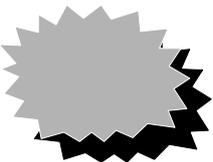
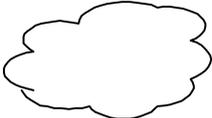
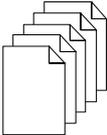
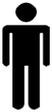
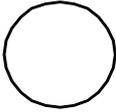
Anexos

Anexo 1 - Símbolos utilizados nas figuras

Ao longo da tese são apresentadas 36 figuras que pretendem ilustrar ideias expostas e a arquitectura do sistema ZACCAR.

Em cada uma delas os símbolos utilizados são muito específicos e a sua repetição é frequente em diferentes figuras (e.g., repositório de dados, sistema de conhecimento). Deste modo, tivemos a preocupação de manter a sua uniformidade de representação ao longo do trabalho.

Nesse sentido e para um melhor esclarecimento do leitor, inserimos neste anexo os símbolos utilizados e respectivo significado.

<i>Símbolo</i>	<i>Significado</i>
	<i>Conhecimento</i>
	<i>Sistema de conhecimento</i>
	<i>Eventos cliente-empresa</i>
	<i>Repositório de dados</i>
	<i>Processos</i>
	<i>Padrões</i>
	<i>Utilizador do conhecimento</i>
	<i>Acções</i>



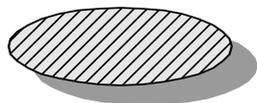
Leitura/registo de eventos



*Processo a detalhar
posteriormente*



*Consumidor/fornecedor do
conhecimento
(homem ou máquina)*



GCCC/Consulta



Metadados



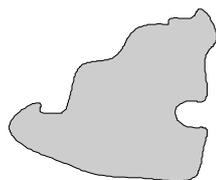
Programa/tabela



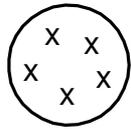
Procedimento



Clientes



Área de intervenção



Conjunto

Anexo 2 – Créditos

- ***Clementine*** é uma marca registada da Integral Solutions Limited.
- ***Cobol*** é uma marca registada da *Fujitsu*.
- ***Foxpro*** é um produto Microsoft.
- ***lex*** é uma marca registada dos *Bell Laboratories*.
- ***Pentium*** é uma marca registada da *Intel*.
- **Win LPA** é uma marca registada da LPA.
- ***Windows NT*** é uma marca registada da *Microsoft*.

Anexo 3 – Alguns módulos dos programas desenvolvidos

Neste anexo encontram-se as listagens com o código *Prolog* dos principais módulos desenvolvidos na construção do protótipo do ZACCAR.

Anexo 3.1 - Programa confronto

```
% programa f_confronto.pl
% Fase de confronto do conhecimento
:-op(850,fx,regra).
:-op(750,xfx,entao).
:-op(720,fx,se).
:-op(710,xfy,e).
:-op(700,xfx,em).
:-op(860,yfx,[evolucao,parametros]).

f_confronto:-
    escreve,
    consult('temp1.pl'),
    consult('bccc.pl'),
    /* open('regras_novas.pl',write),
    close('regras_novas.pl'), */
    /* importa data do sistema */
    ver(_,_D,_),
    confronto(D).

escreve :-
    tell('temp2.pl'),
    write(':-op(860,yfx,evolucao).'),nl,
    write(':-op(850,fx,regra).'),nl,
    write(':-op(750,xfx,entao).'),nl,
    write(':-op(720,fx,se).'),nl,
    write(':-op(710,xfy,e).'),nl,
    write(':-op(700,xfx,em).'),nl,
    told.

confronto(D) :-
    regra se A entao B parametros[C1,C2],
    trata(D,A,B,C1,C2),
    fail.
confronto(_).

trata(D,A,B,C1,C2) :-
    regra se A entao B evolucao Z,
    escreve_regra(A,B,[(data(D),comentario("),parametros(C1,C2))\|Z]),!.

trata(D,A,B,C1,C2) :-
    escreve_regra(A,B,[(data(D),comentario("),parametros(C1,C2))]),!.

escreve_regra(X,Y,T) :-
    tell('temp2.pl'),
    write(regra), write(' '),
    write(se), write(' '), writeq(X),nl,
    write(entao),nl,
    writeq(Y), nl,
    write(evolucao),nl,
```

```
write('['),
writelist(T),
write(']'),
write(' '),nl,
told.
```

```
writelist([]).
writelist([Q|_]):-
    write(' '),writeq(Q),write(' '),nl,
    writelist([]).
writelist([Q|L]) :-
    write(' '),writeq(Q),write(' '),write(' '),nl,
    writelist(L).
```

Anexo 3.2 - Programa *interface*

```

% programa f_interface.pl
% Fase de participação do gestor do conhecimento

:-op(850,fx,regra).
:-op(750,xfx,entao).
:-op(720,fx,se).
:-op(710,xfy,e).
:-op(700,xfx,em).
:-op(860,yfx,evolucao).

f_interface :-
    consult('temp2.pl'),
    escreve,
    actualiza.

escreve:-
    tell('temp3.pl'),
    write(':-op(860,yfx,evolucao).'),nl,
    write(':-op(850,fx,regra).'),nl,
    write(':-op(750,xfx,entao).'),nl,
    write(':-op(720,fx,se).'),nl,
    write(':-op(710,xfy,e).'),nl,
    write(':-op(700,xfx,em).'),nl,
    told.

actualiza :-
    regra se A entao B evolucao Z,
    trata(A,B,Z),
    fail.

actualiza.
% visualizacao de regras para actualizacao pelo gestor do conhecimento
trata(A,B,Z) :-
    write('*****')
),nl,nl,
    write(regra),write(' '),
    write(se),write(' '),
    writeq(A),nl,
    write(entao),nl,
    writeq(B),nl,
    write(evolucao),nl,
    writelist(Z),nl,
    write('*****')
),nl,nl,
    apresenta_menu,
    write('Escolha um dos comandos'),nl,
    write('>'),
    le(Opcao),

```

```

executa_comando(Opcao,A,B,Z),!.

le(X):-
  read(X),X>=0,X<=5.
le(X):-
  write('opcao invalida'),
  le(X).

apresenta_menu :-
  write('Comandos disponiveis
        (introduza o numero 1, 2, 3 ou 4):'),
  nl,
  write('1 - Padrao a eliminar por ja nao ter valor. '), nl,
  write('2 - A grande variacao dos parametros neste periodo justifica-se porque ...'), nl,
  write('3 - A estabilidade dos parametros mostra que nao houve evolucao do padrao'), nl,
  write('4 - Motivos de grande variacao desconhecidos carecendo de analise'),nl,nl,
  write('5 - Outra opcao'), nl.

% actualizacao de regras
executa_comando(1,_,_,_).
%regra eliminada

executa_comando(2,A,B,Z) :-
  write('Introduza o comentario'),nl,
  write('>'),
  read(Novo_coment),
  altera_comentario(A,B,Z,Novo_coment).

executa_comando(3,A,B,Z) :-
  altera_comentario(A,B,Z,'Os padroes continuam estaveis').

executa_comando(4,A,B,Z) :-
  altera_comentario(A,B,Z,'Sem justificacao para alteracoes verificadas').

executa_comando(5,A,B,Z) :-
  write('Introduza o comentario'),nl,
  write('>'),
  read(Novo_coment),
  altera_comentario(A,B,Z,Novo_coment).

altera_comentario(X,Y,[(E1, _E2)\R],N) :-
  escreve_regra(X,Y,[(E1,comentario(N),E2)\R]).

writelist([]).
writelist([Q\[]]):-
  write(' '),writeq(Q),write(' '),nl,
  writelist([]).
writelist([Q\L]) :-
  write(' '),writeq(Q),write(' '),write(' '),nl,
  writelist(L).

escreve_regra(X,Y,T) :-
  tell('temp3.pl'),

```

```
write(regra), write(' '),  
write(se), write(' '), writeq(X),nl,  
write(entao),nl,  
writeq(Y), nl,  
write(evolucao),nl,  
write('['),  
writelist(T),  
write(']'),  
write('.'),nl,  
told.
```

Anexo 3.3 - Programa *assimilação*

```

% programa f_assimilacao.pl
% Fase de assimilação do conhecimento

:-op(850,fx,[regra1,regra]).
:-op(750,xfx,entao).
:-op(720,fx,se).
:-op(710,xfy,e).
:-op(700,xfx,em).
:-op(860,yfx,[evolucao,evolucao1]).

f_assimilacao:-
  consult('change.pl'),
  change('bccc.pl','regra`,`regra1`,`1),
  change('bccc.pl','evolucao`,`evolucao1`,`1),
  consult('bccc.pl'),
  consult('temp3.pl'),
  confronto,
  /* programa batch que se encarrega de copiar temp3.pl+temp4.pl para bccc.pl */
  exec('copia34_bccc.bat',"S).

confronto :-
  regra1 se A entao B evolucao1 W,
  trata(A,B,W),
  fail.
confronto.

trata(A,B,_) :-
  /* se a regra existe nos 2 ficheiros consultados nao faz nada */
  regra se A entao B evolucao _!.

trata(A,B,W) :-
  /* se a regra so existe em bccc entao deve escreve-la em temp4 */
  tell('temp4.pl'),
  write(regra),write(' '),
  write(se), write(' '),writeq(A),nl,
  write(entao),nl,
  writeq(B),nl,
  write(evolucao),nl,
  write(' '),
  writelist(W),
  write(' '),
  write(' '),nl,
  told,!.

writelist([]).
writelist([Q|_]):-
  write(' '),writeq(Q),write(' '),nl,
  writelist(_).

```

```
writelist([Q|L]) :-  
  write(' '),writeq(Q),write(' '),write(', '),nl,  
  writelist(L).
```


Anexo 4 – Inquéritos para obtenção dos dados utilizados no caso 1

Iniciámos o caso 1 com o lançamento de um inquérito em três estabelecimentos de ensino superior (Instituto Superior de Contabilidade e Administração de Coimbra, Departamento de Sistemas de Informação da Universidade do Minho e Faculdade de Ciências e Tecnologia da Universidade de Coimbra) e em quatro momentos distintos. Os dados obtidos com as respostas dos alunos permitiram-nos a elaboração do estudo apresentado em 6.1.

Os inquéritos repetem-se nos quatro momentos à excepção do grupo B em que é proposto, em cada momento, uma modalidade de viagem diferente como se pode observar nos exemplares dos inquéritos que inserimos neste anexo.

Inquérito – 30-10-2000

Suponha que a resposta a este inquérito lhe é solicitada pela empresa *e-tours* da qual faz parte um departamento que se dedica, exclusivamente, à organização de viagens de finalistas. Este departamento pretende obter o parecer do maior número de estudantes tendo em vista uma oferta de uma viagem a Cuba.

A -

Qual a sua situação?

Estudante Trabalhador-estudante

Que curso frequenta? _____

Em que ano está matriculado? _____

Que idade tem? _____

Qual o seu sexo?

Masculino Feminino

Onde reside fora do tempo de aulas?

Freguesia: _____

Concelho: _____

Quantas pessoas compõem o seu agregado familiar? _____

Profissionalmente, qual a situação actual dos seus pais?

	Pai	Mãe
Agricultor	<input type="checkbox"/>	<input type="checkbox"/>
Comerciante	<input type="checkbox"/>	<input type="checkbox"/>
Dirigente/gestor de empresa	<input type="checkbox"/>	<input type="checkbox"/>
Dona(o) de casa	<input type="checkbox"/>	<input type="checkbox"/>
Estudante	<input type="checkbox"/>	<input type="checkbox"/>
Funcionário administrativo	<input type="checkbox"/>	<input type="checkbox"/>
Empregado comercial	<input type="checkbox"/>	<input type="checkbox"/>
Operário	<input type="checkbox"/>	<input type="checkbox"/>
Professor	<input type="checkbox"/>	<input type="checkbox"/>
Profissional liberal	<input type="checkbox"/>	<input type="checkbox"/>
Quadro médio de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Quadro superior de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Técnico especializado	<input type="checkbox"/>	<input type="checkbox"/>
Outro	<input type="checkbox"/>	<input type="checkbox"/>

B - A *e-tours* propõe uma viagem a Cuba com:

- Estadia de 8 dias/7 noites na praia de Varadero
- Instalações em hotel de 4 estrelas em regime de TI (Tudo Incluído)
- Preço por pessoa em quarto duplo: 900 euros

Participaria neste programa?

Sim Não

Muito obrigada pelo tempo despendido a preencher este inquérito

Inquérito – 30-11-2000

Suponha que a resposta a este inquérito lhe é solicitada pela empresa *e-tours* da qual faz parte um departamento que se dedica, exclusivamente, à organização de viagens de finalistas. Este departamento pretende obter o parecer do maior número de estudantes tendo em vista uma oferta de uma viagem a Cuba.

A -

Qual a sua situação?

Estudante Trabalhador-estudante

Que curso frequenta? _____

Em que ano está matriculado? _____

Que idade tem? _____

Qual o seu sexo?

Masculino Feminino

Onde reside fora do tempo de aulas?

Freguesia: _____

Concelho: _____

Quantas pessoas compõem o seu agregado familiar? _____

Profissionalmente, qual a situação actual dos seus pais?

	Pai	Mãe
Agricultor	<input type="checkbox"/>	<input type="checkbox"/>
Comerciante	<input type="checkbox"/>	<input type="checkbox"/>
Dirigente/gestor de empresa	<input type="checkbox"/>	<input type="checkbox"/>
Dona(o) de casa	<input type="checkbox"/>	<input type="checkbox"/>
Estudante	<input type="checkbox"/>	<input type="checkbox"/>
Funcionário administrativo	<input type="checkbox"/>	<input type="checkbox"/>
Empregado comercial	<input type="checkbox"/>	<input type="checkbox"/>
Operário	<input type="checkbox"/>	<input type="checkbox"/>
Professor	<input type="checkbox"/>	<input type="checkbox"/>
Profissional liberal	<input type="checkbox"/>	<input type="checkbox"/>
Quadro médio de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Quadro superior de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Técnico especializado	<input type="checkbox"/>	<input type="checkbox"/>
Outro	<input type="checkbox"/>	<input type="checkbox"/>

B - A *e-tours* propõe uma viagem a Cuba com:

- Circuito histórico, com a duração de um dia, na cidade de Havana
- Estadia de 7 dias/6 noites na praia de Varadero
- Instalações em hotel de 4 estrelas em regime de TI (Tudo Incluído)
- Preço por pessoa em quarto duplo: 900 euros

Participaria neste programa?

Sim Não

Muito obrigada pelo tempo despendido a preencher este inquérito

Inquérito – 30-12-2000

Suponha que a resposta a este inquérito lhe é solicitada pela empresa *e-tours* da qual faz parte um departamento que se dedica, exclusivamente, à organização de viagens de finalistas. Este departamento pretende obter o parecer do maior número de estudantes tendo em vista uma oferta de uma viagem a Cuba.

A -

Qual a sua situação?

Estudante Trabalhador-estudante

Que curso frequenta? _____

Em que ano está matriculado? _____

Que idade tem? _____

Qual o seu sexo?

Masculino Feminino

Onde reside fora do tempo de aulas?

Freguesia: _____

Concelho: _____

Quantas pessoas compõem o seu agregado familiar? _____

Profissionalmente, qual a situação actual dos seus pais?

	Pai	Mãe
Agricultor	<input type="checkbox"/>	<input type="checkbox"/>
Comerciante	<input type="checkbox"/>	<input type="checkbox"/>
Dirigente/gestor de empresa	<input type="checkbox"/>	<input type="checkbox"/>
Dona(o) de casa	<input type="checkbox"/>	<input type="checkbox"/>
Estudante	<input type="checkbox"/>	<input type="checkbox"/>
Funcionário administrativo	<input type="checkbox"/>	<input type="checkbox"/>
Empregado comercial	<input type="checkbox"/>	<input type="checkbox"/>
Operário	<input type="checkbox"/>	<input type="checkbox"/>
Professor	<input type="checkbox"/>	<input type="checkbox"/>
Profissional liberal	<input type="checkbox"/>	<input type="checkbox"/>
Quadro médio de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Quadro superior de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Técnico especializado	<input type="checkbox"/>	<input type="checkbox"/>
Outro	<input type="checkbox"/>	<input type="checkbox"/>

B - A *e-tours* propõe uma viagem a Cuba com:

- Estadia de 7 dias/6 noites na praia de Varadero
- Instalações em hotel de 4 estrelas em regime de TI (tudo incluído)
- Preço por pessoa em quarto duplo: 900 euros, pago em 3 prestações.

Participaria neste programa?

Sim Não

Muito obrigada pelo tempo despendido a preencher este inquérito

Inquérito – 30-01-2001

Suponha que a resposta a este inquérito lhe é solicitada pela empresa *e-tours* da qual faz parte um departamento que se dedica, exclusivamente, à organização de viagens de finalistas. Este departamento pretende obter o parecer do maior número de estudantes tendo em vista uma oferta de uma viagem a Cuba.

A -

Qual a sua situação?

Estudante Trabalhador-estudante

Que curso frequenta? _____

Em que ano está matriculado? _____

Que idade tem? _____

Qual o seu sexo?

Masculino Feminino

Onde reside fora do tempo de aulas?

Freguesia: _____

Concelho: _____

Quantas pessoas compõem o seu agregado familiar? _____

Profissionalmente, qual a situação actual dos seus pais?

	Pai	Mãe
Agricultor	<input type="checkbox"/>	<input type="checkbox"/>
Comerciante	<input type="checkbox"/>	<input type="checkbox"/>
Dirigente/gestor de empresa	<input type="checkbox"/>	<input type="checkbox"/>
Dona(o) de casa	<input type="checkbox"/>	<input type="checkbox"/>
Estudante	<input type="checkbox"/>	<input type="checkbox"/>
Funcionário administrativo	<input type="checkbox"/>	<input type="checkbox"/>
Empregado comercial	<input type="checkbox"/>	<input type="checkbox"/>
Operário	<input type="checkbox"/>	<input type="checkbox"/>
Professor	<input type="checkbox"/>	<input type="checkbox"/>
Profissional liberal	<input type="checkbox"/>	<input type="checkbox"/>
Quadro médio de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Quadro superior de empresa/função pública	<input type="checkbox"/>	<input type="checkbox"/>
Técnico especializado	<input type="checkbox"/>	<input type="checkbox"/>
Outro	<input type="checkbox"/>	<input type="checkbox"/>

B - A *e-tours* propõe uma viagem a Cuba com:

- Circuito histórico, com a duração de um dia, na cidade de Havana
- Estadia de 6 dias/5 noites na praia de Varadero
- Instalações em hotel de 4 estrelas em regime de TI (tudo incluído)
- Preço por pessoa em quarto duplo: 900 euros, pago em 3 prestações.

Participaria neste programa?

Sim Não

Muito obrigada pelo tempo despendido a preencher este inquérito

Anexo 5 – Extracto da base de dados utilizada no caso 1

As respostas obtidas nos inquéritos lançados numa população estudantil, matriculada em vários estabelecimentos do ensino superior, foram registadas numa base de dados convenientemente estruturada tendo em atenção os objectivos do estudo que se pretendia realizar.

Apresentamos, neste anexo, um extracto dessa base de dados.

<i>Data</i>	<i>Situacao</i>	<i>Curso</i>	<i>Ano</i>	<i>Idade</i>	<i>Sexo</i>	<i>Tipo_res</i>	<i>Zona</i>	<i>Quantos</i>	<i>Prof_pai</i>	<i>Prof_mae</i>	<i>Opcao</i>
30-10-2000	E	2	2	20	M	R	C	5	1	4	N
30-10-2000	T	3	2	24	F	R	C	3	3	3	N
30-10-2000	E	2	2	21	F	R	C	4	8	14	N
30-10-2000	E	2	2	21	F	R	C	4	1	4	N
30-10-2000	T	2	2	21	M	R	C	3	2	2	N
30-10-2000	E	2	2	19	M	R	C	5	7	4	N
30-10-2000	E	2	2	19	F	R	C	4	14	4	S
30-10-2000	E	2	2	19	F	R	C	4	2	9	S
30-10-2000	E	2	2	22	M	R	O	0	14	14	S
30-10-2000	E	2	2	19	F	R	C	4	8	8	S
30-10-2000	E	2	2	19	F	R	C	4	2	14	N
30-10-2000	E	2	2	21	F	U	C	3	6	11	S
30-10-2000	E	2	2	21	M	U	C	2	6	4	S
30-10-2000	E	2	2	43	M	U	C	4	2	3	S
30-10-2000	E	2	2	21	F	R	C	4	3	14	S
30-10-2000	E	2	2	20	F	U	C	3	3	9	N
30-10-2000	E	2	2	21	M	U	C	5	10	4	N
30-10-2000	T	2	2	22	M	R	C	4	11	6	S
30-10-2000	E	2	2	23	M	R	C	4	2	2	S
30-10-2000	E	2	2	19	M	R	C	5	13	9	N
30-10-2000	E	2	2	19	M	R	C	3	8	4	N
30-10-2000	E	2	2	22	M	U	N	4	2	14	N
30-10-2000	E	2	2	20	F	R	C	6	2	4	N
30-10-2000	E	2	2	20	F	U	C	3	2	11	N
30-10-2000	E	1	4	21	F	R	C	3	14	14	N
30-10-2000	E	1	1	19	F	R	S	3	2	4	S
30-10-2000	E	1	4	24	F	R	C	4	11	11	N
30-10-2000	E	1	2	22	M	U	C	4	14	4	N
30-10-2000	T	1	1	29	F	R	C	3	3	4	N
30-10-2000	E	1	3	21	F	R	C	3	1	1	N
30-10-2000	E	1	3	20	F	R	C	4	2	2	S
30-10-2000	E	1	3	24	F	R	C	3	14	6	N
30-10-2000	E	2	1	19	M	U	C	0	9	11	N
30-10-2000	E	2	1	23	M	U	C	4	10	12	N
30-10-2000	T	2	2	22	F	R	C	3	2	2	N
30-10-2000	E	1	2	22	F	R	O	4	3	3	S
30-10-2000	E	1	2	23	F	R	C	3	1	1	N
30-10-2000	E	1	3	21	M	U	C	4	8	14	N
30-10-2000	E	1	4	23	M	U	C	3	14	7	S
30-10-2000	T	1	4	24	F	U	C	5	2	2	N
30-10-2000	T	1	2	23	M	U	C	3	12	11	N
30-10-2000	T	1	1	24	F	R	C	4	10	1	S
30-10-2000	E	1	2	23	F	U	C	4	2	2	N
30-10-2000	T	3	1	23	M	U	S	0	3	12	S
30-10-2000	E	1	4	24	M	R	O	5	11	6	N
30-10-2000	T	1	2	27	M	R	C	4	2	4	N

30-10-2000	E	1	2	26	F	R	C	5	2	2	N
30-10-2000	T	1	2	21	M	U	C	3	12	12	S
30-10-2000	E	1	2	21	M	R	C	4	14	4	S
30-10-2000	E	1	2	22	F	R	N	4	11	4	S
30-10-2000	E	1	2	21	F	R	C	5	10	4	S
30-10-2000	E	1	2	23	F	R	C	5	2	2	N

Anexo 6 – Regras obtidas, no caso 1, com o *Clementine*

A aplicação do algoritmo GRI do *Clementine* nos dados contidos na base de dados da agência de viagens fictícia *e-tours*, em quatro momentos diferentes, conduziu à obtenção de um conjunto de padrões de comportamento traduzidos pelas regras que se anexam.

Anexo 6.1 Regras obtidas no momento 1

Opção == N <= Agregado_familiar < 3.5 & Curso == 1 & Profissao_pai == -- (18:3.529%, 1.0)
Opção == N <= Agregado_familiar < 3.5 & Idade > 23.5 & Origem == R (62:12.157%, 0.87)
Opção == N <= Ano == 2 & Idade > 23.5 & Idade < 28.5 (30:5.882%, 1.0)
Opção == N <= Ano == 3 & Profissao_pai == -- (15:2.941%, 1.0)
Opção == N <= Curso == 1 & Agregado_familiar < 3.5 & Profissao_pai == -- (18:3.529%, 1.0)
Opção == N <= Curso == 1 & Ano == 3 & Profissao_pai == -- (14:2.745%, 1.0)
Opção == N <= Curso == 1 & Idade > 23.5 & Profissao_pai == -- (17:3.333%, 1.0)
Opção == N <= Curso == 1 & Origem == R & Profissao_pai == -- (18:3.529%, 1.0)
Opção == N <= Curso == 1 & Profissao_pai == -- (24:4.706%, 1.0)
Opção == N <= Curso == 1 & Area_residencia == C & Profissao_pai == -- (22:4.314%, 1.0)
Opção == N <= Curso == 1 & Sexo == F & Profissao_pai == -- (16:3.137%, 1.0)
Opção == N <= Curso == 2 & Ano > 3 & Profissao_pai == 14 (16:3.137%, 0.81)
Opção == S <= Ano == 2 & Idade > 31.5 & Agregado_familiar > 3.5 (5:0.98%, 1.0)
Opção == S <= Curso == 3 & Agregado_familiar > 3.5 & Agregado_familiar < 4.5 (5:0.98%, 1.0)
Opção == S <= Curso == 3 & Origem == R & Profissao_mae == 14 (4:0.784%, 1.0)
Opção == S <= Curso == 3 & Profissao_mae == 14 (5:0.98%, 1.0)
Opção == S <= Curso == 3 & Area_residencia == C & Profissao_mae == 14 (5:0.98%, 1.0)
Opção == S <= Curso == 3 & Sexo == F & Profissao_mae == 14 (4:0.784%, 1.0)
Opção == S <= Curso == 5 & Idade > 22.5 (18:3.529%, 0.67)
Opção == S <= Curso == 5 (43:8.431%, 0.53)
Opção == S <= Situacao == E & Curso == 5 & Idade > 22.5 (15:2.941%, 0.73)
Opção == S <= Idade < 19..5 & Area_residencia == C & Agregado_familiar > 3. 5 (23:4.51%, 0.35)

Anexo 6.2 Regras obtidas no momento 2

Opcao == N <= Ano == 2 & Agregado_familiar < 2.5 & Idade < 30.5 (12:3.361%, 1.0)
Opcao == N <= Ano == 2 & Profissao_pai == 14 & Profissao_mae == 14 (12:3.361%, 1.0)
Opcao == N <= Curso == 2 & Ano > 3 & Profissao_pai == 14 (16:3.137%, 0.81)
Opcao == N <= Situacao == T & Curso == 1 & Profissao_pai == 8 (12:3.361%, 1.0)
Opcao == N <= Situacao == T & Curso == 2 & Sexo == F (13:3.641%, 1.0)
Opcao == N <= Situacao == T & Idade < 23.5 & Sexo == F (30:8.403%, 0.93)
Opcao == N <= Situacao == T & Profissao_pai == 8 (13:3.641%, 1.0)
Opcao == N <= Situacao == T & Area_residencia == C & Profissao_pai == 8 (13:3.641%, 1.0)
Opcao == S <= Agregado_familiar > 2.5 & Curso == 2 & Idade < 19.5 (10:2.801%, 0.8)
Opcao == S <= Agregado_familiar > 2.5 & Idade < 19.5 & Curso == 2 (10:2.801%, 0.8)
Opcao == S <= Agregado_familiar > 2.5 & Area_residencia == N & Profissao_pai == 2 (4:1.12%, 1.0)
Opcao == S <= Ano == 3 & Agregado_familiar < 1.5 & Origem == U (6:1.681%, 1.0)
Opcao == S <= Ano == 3 & Origem == U & Agregado_familiar < 1.5 (6:1.681%, 1.0)
Opcao == S <= Curso == 1 & Idade > 11.5 & Profissao_pai == 10 (8:2.241%, 0.88)
Opcao == S <= Curso == 1 & Profissao_pai == 10 (8:2.241%, 0.88)
Opcao == S <= Situacao == E & Curso == 5 & Idade > 22.5 (4:1%, 0.50)
Opcao == S <= Curso == 2 & Agregado_familiar > 2.5 & Idade < 19.5 (10:2.801%, 0.8)
Opcao == S <= Idade < 19.5 & Agregado_familiar > 3.5 & Curso == 2 (9:2.521%, 0.89)
Opcao == S <= Idade < 19.5 & Area_residencia == C & Agregado_familiar > 3.5 (15:4.202%, 0.73)
Opcao == S <= Idade < 19.5 & Situacao == E & Sexo == F (12:3.361%, 0.75)
Opcao == S <= Profissao_pai == 10 (14:3.922%, 0.71)
Opcao == S <= Sexo == M & Profissao_pai == 2 (10:2.801%, 0.8)

Anexo 6.3 Regras obtidas no momento 3

Opção == N <= Ano == 1 & Idade > 23.5 (38:7.451%, 1.0)
Opção == N <= Ano == 2 & Idade > 24.5 (36:7.059%, 1.0)
Opção == N <= Curso == 1 & Ano == 2 & Idade > 22.5 (53:10.392%, 1.0)
Opção == N <= Curso == 1 & Idade > 29.5 (37:7.255%, 1.0)
Opção == N <= Curso == 2 & Ano > 3 & Profissao_pai == 14 (13:5.349%, 0.77)
Opção == N <= Idade > 29.5 (45:8.824%, 1.0)
Opção == N <= Situação == E & Agregado_familiar > 5.5 & Sexo == F (6:1.176%, 0.5)
Opção == N <= Situação == E & Origem == U & Profissao_mae == 4 (36:7.059%, 1.0)
Opção == N <= Situação == E & Sexo == F & Agregado_familiar > 5.5 (6:1.176%, 0.5)
Opção == N <= Situação == T & Ano == 1 & Idade > 23.5 (37:7.255%, 1.0)
Opção == N <= Situação == T & Curso == 1 & Idade < 22.5 (29:5.686%, 0.76)
Opção == N <= Situação == T & Idade > 29.5 (42:8.235%, 1.0)
Opção == S <= Ano == 2 & Origem == R & Profissao_mae == 11 (3:0.588%, 0.67)
Opção == S <= Ano == 3 & Profissao_pai == 7 & Profissao_mae == 7 (2:0.392%, 1.0)
Opção == S <= Ano == 3 & Sexo == M & Profissao_pai == 7 (3:0.588%, 0.67)
Opção == S <= Profissao_pai == 7 & Profissao_mae == 7 (2:0.392%, 1.0)
Opção == S <= Area_residencia == C & Profissao_pai == 7 & Profissao_mae == 7 (2:0.392%, 1.0)
Opção == S <= Sexo == F & Origem == U & Profissao_mae == 3 (2:0.392%, 1.0)
Opção == S <= Situação == E & Profissao_pai == -- & Profissao_mae == 4 (3:0.588%, 0.67)
Opção == S <= Situação == E & Curso == 5 & Idade > 22.5 (4:2%, 0.50)
Opção == S <= Situação == E & Profissao_pai == 7 & Profissao_mae == 7 (2:0.392%, 1.0)
Opção == S <= Situação == E & Sexo == F & Agregado_familiar > 5.5 (6:1.176%, 0.5)
Opção == S <= Idade < 19.5 & Area_residencia == C & Agregado_familiar > 3.5 (4:1.65%, 0.25)

Anexo 6.4 Regras obtidas no momento 4

Opção == N <= Ano == 2 & Sexo == F & Profissao_mae == 2 (7:1.373%, 0.14)
Opção == N <= Curso == 1 & Idade > 28.5 & Agregado_familiar < 2.5 (21:4.118%, 1.0)
Opção == N <= Curso == 2 & Ano == 2 & Idade < 20.5 (23:4.51%, 1.0)
Opção == N <= Curso == 2 & Ano > 3 & Profissao_pai == 14 (10:4.975%, 0.50)
Opção == N <= Curso == 5 & Idade > 22.5 (18:3.529%, 1.0)
Opção == N <= Curso == 5 & Tipo_residencia == R (31:6.078%, 0.97)
Opção == N <= Idade > 28.5 & Agregado_familiar < 2.5 & Curso == 1 (21:4.118%, 1.0)
Opção == N <= Idade > 28.5 & Curso == 1 & Agregado_familiar < 2.5 (21:4.118%, 1.0)
Opção == N <= Idade > 28.5 & Situacao == T & Ano == 3 (19:3.725%, 1.0)
Opção == N <= Sexo == F & Tipo_residencia == R & Profissao_mae == 2 (16:3.137%, 0.44)
Opção == N <= Situacao == E & Ano == 2 & Idade > 22.5 (25:4.902%, 0.52)
Opção == N <= Situacao == E & Curso == 2 & Idade < 20.5 (28:5.49%, 1.0)
Opção == N <= Situacao == T & Ano == 3 & Idade > 28.5 (19:3.725%, 1.0)
Opção == N <= Situacao == T & Idade > 26.5 & Idade < 36.5 (50:9.804%, 0.94)
Opção == N <= Situacao == T & Tipo_residencia == U & Agregado_familiar < 2.5 (20:3.922%, 1.0)
Opção == S <= Ano == 2 & Sexo == F & Profissao_mae == 2 (7:1.373%, 0.86)
Opção == S <= Curso == 1 & Ano == 1 & Area_residencia == N (3:0.588%, 1.0)
Opção == S <= Curso == 4 (4:0.784%, 1.0)
Opção == S <= Idade < 19.5 & Area_residencia == C & Agregado_familiar > 3.5 (34:1.49%, 0.67)
Opção == S <= Situacao == E & Ano == 1 & Area_residencia == N (3:0.588%, 1.0)
Opção == S <= Situacao == E & Curso == 4 (3:0.588%, 1.0)
Opção == S <= Situacao == T & Idade > 26.5 & Idade < 36.5 (50:9.804%, 0.06)

Anexo 7 – Extracto da base de dados usada no caso 2

Neste anexo apresentamos um extracto da base de dados utilizada no caso 2.

Por razões de confidencialidade, dado que neste caso os dados são reais, não podemos apresentar nem todos os campos nem todas as ocorrências da base de dados. Assim, o extracto apresentado apenas tem apenas a intenção de permitir ao leitor fazer uma ideia do tipo e estrutura dos dados contidos na base de dados da AMMA.

Anexos

<i>N-encomenda</i>	<i>N-Linha</i>	<i>N_Cliente</i>	<i>Localidade</i>	<i>N_Modelo</i>	<i>Tipo</i>	<i>Variantes</i>	<i>Referencia</i>	<i>Data_entrada</i>	<i>Data_pedido</i>	<i>Total</i>
990335	9	2110024	INGLATERRA	22453024	fato	casaco c/2 aberturas,calça c/viola	14301.89R	19990825	19990420	4
990481	5	2110557	ESPANHA	40052002	calça	calça sem acabamentos	000TT.519A	19990826	19990526	6
990481	4	2110557	ESPANHA	40052002	calça	calça sem acabamentos	000TT.540A	19990826	19990526	6
990559	1	2110412	FRANÇA	40055001	calça	calça sem acabamentos	96032.03 A	19990702	19990615	100
992593	1	2110557	ESPANHA	21058202	fato	casaco c/2 aberturas,calça s/acabamentos	10492.97 C	19990826	19990621	1
992591	1	2110557	ESPANHA	11000202	casaco	casaco com 2 aberturas	05050.05 C	19990826	19990621	1
992583	1	2110007	AMADORA	58006000	colete	colete	13557.89 C	19990628	19990617	1
990563	15	2110101	LISBOA	21753200	fato	casaco c/2 aberturas,calça s/acabamentos	14167.88 C	19991025	19990702	11
990285	4	2110105	ALMADA	31253220	fato com colete	casaco c/2 aberturas,calça c/viola	14555.89 C	19991110	19990304	5
992595	1	2110113	ÉVORA	40010020	calça	calça com virola	11679.83 I	19990630	19990622	1
992596	1	2110129	S.PEDRO AVIOSO	21710220	fato	casaco c/2 aberturas,calça c/viola	00120.94 C	19990707	19990622	1
990345	10	2110324	IRLANDA	21753004	fato	casaco s/aberturas,calça s/acabamentos	14046.89 R	19990825	19990420	6
990505	4	2110588	INGLATERRA	14900104	casaco	casaco com 1 abertura	78441.980S	19990701	19990527	15
990505	3	2110588	INGLATERRA	14900104	casaco	casaco com 1 abertura	78441.590L	19990701	19990527	11
990505	2	2110588	INGLATERRA	14900104	casaco	casaco com 1 abertura	78441.590R	19990701	19990527	49
990505	1	2110588	INGLATERRA	14900104	casaco	casaco com 1 abertura	78441.590S	19990701	19990527	7
990867	10	2110444	FUNCHAL	40052020	calça	calça com virola	15343.73 A	20000201	19990922	7
990867	11	2110444	FUNCHAL	40052020	calça	calça com virola	15343.75 A	20000201	19990922	7
990868	3	2110320	MEM MARTINS	21753020	fato	casaco c/2 aberturas,calça c/viola	15265.92 C	20000215	19990922	5
990868	2	2110320	MEM MARTINS	31753020	fato com colete	casaco s/aberturas,calça c/viola	15280.98 C	20000215	19990922	5

Índice de Autores

	Dougherty, 1994	117, 118
	E	
	Etzioni, 1980	1
	F	
	Fayyad <i>et al.</i> , 1996	1, 50
	Fayyad <i>et al.</i> , 1996a	73, 74, 105
	Fayyad <i>et al.</i> , 1996b	73, 80, 81, 107
	Fayyad <i>et al.</i> , 1996c	4, 50
	Fayyad, 1996	52, 71
	Feldens, 2000	46
	Ferrão, 2001	53
	Figher, 1987	74
	Franks, 1997	85
	Frawley <i>et al.</i> , 1991	4, 50, 71, 81
	Freitas & Lavington, 1998	68, 101
	Frost, 1987	111
	Furtado, 2000	59
	G	
	Garshol, 2000	113
	Groth, 2000	69, 75
	Gulliksen & Sadblad, 1995	185
	Gulliksen, 1996	185
	Gutiérrez <i>et al.</i> , 2000	52
	H	
	Han & Kamber, 2001	1, 77, 81
	<i>Hand</i> , 1994	137
	Hewett <i>et al.</i> , 1997	185
	Holsheimer & Kersten, 1994	87
	Holsheimer, 1995	71, 78, 82
	Hong & Mao, 1991	107
	Howe, 1995	113
	Hu, 1995	75
	I	
	Inmon & Hackathorn, 1994	63, 65
	Inmon, 1992	62
	Inmon, 1996	75
	K	
	Kenan Technologies, 1995	59
	Klösgen & Zytkow, 1996	73
	Kohavi & Provost, 2000	35, 96
	Krogh <i>et al.</i> , 1999	38
	Kudyba & Hoptroff, 2001	4, 81, 186
	L	
	Lopes <i>et al.</i> , 1999	42
	Lucas, 1997	107
	Lugger & Stubblefield, 1998	118
A		
Agrawal <i>et al.</i> , 1993	1, 84, 88	
Aho <i>et al.</i> , 1985	111	
Anand & Kahn, 1993	78	
B		
Baião, 2000	51	
Berry & Linoff, 1997	69, 107	
Berry & Linoff, 2000	4, 18, 27, 52, 76, 83	
Berson <i>et al.</i> , 1999	4, 8, 16, 18, 25, 31, 69, 71, 76, 82	
Bits & Pixels, 2001	52, 53	
Blackwell <i>et al.</i> , 2001	30	
Brachman & Anand, 1996	4, 50, 71	
Brachman & Levesque, 1985	111	
Brachman <i>et al.</i> , 1996	75, 78, 81	
Brackett, 1999	3, 81	
Brathwaite, 1992	54	
Bratko, 1990	111, 112, 117, 118	
Brito, 2000	32	
Brodie & Mylopoulos, 1986	111	
Brown, 2000	4, 16, 18, 28, 36, 52	
Burnett, 2001	4, 16, 18	
C		
Cabena <i>et al.</i> , 1997	38, 62, 71, 72, 76, 79, 82	
Cabibbo & Torlone, 1997	59	
Cardoso, 1995	117	
Carvalho & Morais, 2001	2, 12, 42, 45	
Cassell & Symon, 1997	12	
Cavaye, 1996	12, 13	
Choo, 1996	43	
<i>Clementine</i> , 1997	8, 82, 138, 139, 162	
Clocksinn & Mellish, 1987	118	
Codd, 1970	56	
Coelho & Cotta, 1988	118	
Curry & Curry, 2000	20, 22, 51	
D		
Date, 1995	55, 58	
Demarest, 1994	69	
Dhar & Stein, 1997	4, 84, 112	

M	
Mannila & Toivonen, 1996	86
Marciniak, 1994	118, 185
Mason & Brown, 1991	111
Mello, 1997	75
Montcel, 1972	16
Moriarty, 1999	103
N	
Newell, 2000	3, 4, 16, 18, 22, 34, 52
Newell, 2000a	28
Nonaka & Takeuchi, 1995	38, 43
O	
Okell, 1999	82
OLAP, 1997	59
P	
Parsaye & Chignell, 1993	68, 71, 82, 107
Parsaye, 1997	59
Peppers & Rogers, 2000	19, 21
Pereira, 1998	54, 55, 56
Piatetsky-Shapiro & Frawley, 1991	4,71
Prusak, 1997	42
R	
Rodrigues <i>et al.</i> , 2000	155
S	
Santos, 1999	8, 116, 117, 118
Schafer <i>et al.</i> , 2001	39, 186
Scott & Nicolson, 1991	118
Shindler, 1997	79
Shinghal, 1992	118
Shoam, 1994	118
Silberschatz <i>et al.</i> , 1995	77
Silberschatz <i>et al.</i> , 1996	25, 54, 58, 84, 105
Silberschatz & Tuzhilin., 1996	105
Simoudis & Kerber, 1996	77
Simoudis, 1996	78, 81, 88, 114
Stefix, 1995	1, 5, 6, 51, 111
Sterling & Shapiro, 1994	117
Stonebraker <i>et al.</i> , 1990	55
T	
Thierauf, 1997	59
Toleman, 1996	185
U	
Ullman, 1982	56, 58
Uthurusamy, 1996	77, 81
V	
Vogel, 1996	90
W	
Wand & Wang, 1996	67
Wand <i>et al.</i> , 1995	7, 9
Z	
Zorrinho, 1991	1

Índice Remissivo

A

acções na organização 67, 95
 actividade
 organizacional.....2, 5, 14, 100, 108, 109, 110, 137
 actualização dinâmica 91, 181
 agentes organizacionais..... 42, 74, 96
 álgebra relacional 56, 58
 alterabilidade..... 112
 ambiente real..... 184
 AMMA 155, 158, 182
 bases de dados 178, 231
 analisador lexical 123
 análise da evolução 90, 94, 127, 182
 antecedentes 88, 124, 125, 176, 185
 aprendizagem
 automática..... 71, 73, 77, 98
 algoritmos..... 73
 técnicas 71
 incremental 73, 98, 107
 supervisionada 73
 arquitectura funcional 119, 137
 array multidimensional..... 60
 árvore de decisão..... 82, 83, 89
 assimilador 176, 185
 associações de produtos 23
 ATM ix, 14, 52, 95
 atributo 7, 56, 83, 87, 135, 138, 140, 141, 142, 145,
 147, 149, 154, 160, 161
 derivado 71, 75, 162
 irrelevante 74
 aumento da oferta..... 39
automatic transaction machines..... Ver ATM

B

backward chaining..... 8, 118
 base de conhecimento5, 6, 7, 9, 10, 11, 14, 50, 52, 90,
 91, 98, 103, 106, 109, 112, 120, 126, 133, 152,
 154, 176, 181, 184, 185, 186
 consulta..... 153, 182
 criação..... 93, 101
 do comportamento dos clientes.. ix, x, 22, 97, 101,
 103, 105, 109, 116, 119, 122, 124, 130, 132,

133, 134, 135, 144, 146, 150, 165, 166, 168,
 169, 172, 173, 174, 181, 183, 186
 actualização 102, 103, 132
 consulta 151
 conteúdo 148
 criação 102
 manutenção..... 98
 requisitos 101, 177
 bases de dados... 12, 14, 25, 29, 41, 44, 47, 49, 53, 55,
 56, 58, 59, 60, 61, 62, 63, 68, 69, 70, 72, 74, 77,
 79, 81, 100, 152, 158, 161, 180
 científicas..... 59
 descoberta de conhecimento .. ix, 4, 50, 70, 71, 75,
 77, 96
 definição 71
 função 77
 passos principais..... 72
 peritos..... 102
 processo..... 102
 distribuídas 59
 estatísticas..... 59
 modelos 55
 multidimensionais..... 59
 operacionais 101
 qualidade da descoberta de conhecimento 75
 relacionais..... 56, 59
 sistema de gestão ix, x, 54, 62, 67, 161
 tecnologia 54, 55, 68, 70
 temporais 84
 transaccionais .. 67, 75, 94, 96, 100, 102, 103, 121,
 155, 182
 BCCC..... Ver base de conhecimento do
 comportamento dos clientes
 BI..... Ver *business intelligence*
business intelligence ix, 2, 3, 4, 18, 81

C

C5.0..... 83, 162
call centers..... 14, 19, 24, 52, 95
 caminho de consolidação 107
 campanha de mercado individualizada..... 41
 campanhas. 23, 24, 30, 40, 76, 97, 103, 127, 157, 165,
 167, 171, 173, 174
 campos 56, 60, 61, 71, 78, 160, 161, 231
 cartão

crédito	25, 27, 30, 34, 52, 78
fidelidade	21, 27, 34, 41, 52
magnético.....	50, 52
multibanco	52
caso 1	13, 139, 215, 221, 225
caso 2	155, 231
CCC	Ver conhecimento:do comportamento dos clientes
CHAID	83
ciclo de vida dos produtos.....	39
classificações.....	45
cláusulas de Horn	8, 118
<i>Clementine</i> ..	8, 82, 111, 138, 139, 140, 143, 145, 148, 150, 162, 163, 203, 225
cliente.....	192
alteração do comportamento	47
análise do comportamento de compras	29
aquisição	3, 23, 30
atrair.....	4, 16, 22
atritos	23
canais preferenciais.....	31
ciclo de vida.....	28, 30, 31
comportamento	3, 19, 28, 31, 78, 79, 90, 92, 94, 95, 97, 101, 103, 104, 107, 115, 138, 174, 186
conservar.....	4, 16
diferenciar	19, 32
efectivo	29
evento	24
eventos	28
relacionados.....	23
evolução da satisfação	31
evolução do comportamento.....	3, 93, 94, 115
fases de relacionamento	28
fidelidade	26, 34, 157
fidelizar	3, 19, 22, 32
grau de satisfação.....	23
identificar.....	19, 22, 26, 32
informação	26
interagir.....	19
necessidades e aspirações	17
necessidades individuais	78
oferta à medida das necessidades.....	33
perfil	32
compras	78
personalizar.....	19
ponto de contacto.....	10
potencial	24, 29, 52
processos de relacionamento.....	18
relação estável.....	25
relacionamento.....	10, 17, 28
retenção.....	16, 23, 40
satisfação	19, 23, 31, 32
segmentar.....	22, 29
transacções.....	34, 37, 40, 50, 101
tratamento individualizado.....	23, 25
uma melhor perspectiva.....	33
valor	
do negócio	21
para a empresa	17
cliente-empresa	
ciclo de vida.....	28
contacto	17, 53, 156, 181
eventos.....	14, 48, 50, 91
registos	48
relação	3, 18, 26, 37
relacionamento.....	33, 35, 38
<i>Cobol</i>	158, 203
código de barras	20, 50, 51
comércio electrónico.....	19, 29, 35, 38, 39, 52, 186
compras	
épocas preferencias.....	23
preferências.....	26, 53
conflitos	44, 104, 120, 175, 181
confronto.....	105, 132, 148, 150
processo	107
conhecimento	
aquisição	5, 49, 121
assimilação	121, 130
coerente	183
conceptual.....	42, 43, 47, 50, 181
aquisição.....	77
armazenamento.....	90
gestão	44
confronto	17, 121, 124
consolidação	9, 183
cultural.....	43
do comportamento dos clientes.....	ix, 3, 31, 32, 44, 73, 110, 137, 181
actualizar	153, 182
alimentar.....	100
aquisição.....	96
armazenar	98
benefícios	33
exploração	102
gestão	4, 5, 9, 14
integração automática.....	97
obtenção	4, 39
percurso	44
processo de aquisição, registo e exploração .	94
recurso estratégico.....	37
zelar	5, 9, 14, 93, 108, 110, 181
documentação	9
domínio.....	107
explícito	43
extração.....	82
factual	42, 49
formalismo de representação	91, 104, 111, 122
formulação	5
fornecedor.....	6
gestor	106, 115, 127, 132, 133
importância da formalização.....	47
integração	9, 115
objecto	116
organizacional.....	2, 5, 9, 42, 47, 91, 96, 100, 101, 120, 136, 181
classificação	42
gestor.....	9, 106, 132
participação do gestor	121, 127

processo de consolidação 131, 155, 174, 178, 181, 182
 viabilidade 184
 registo 49, 117
 representação 102, 112, 116, 122
 tipo..... 8
 tácito 43
 transformação 42
 validação 9
 conjunto alvo..... 107
 conversões de tipos de dados 75
cookie 53
 correlações 73
Coverstory Spotlight 78
 CRM ix, 3, 4, 9, 11, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 36, 47, 48, 92, 100, 103, 153, 180, 182, 184, 187
 actividades 47, 153, 182
 estratégia 4, 26, 48
 perspectiva 4
customer relationship management..... Ver CRM

D

dados
 analisar 158, 161
 compreender 158, 161
 inconsistência..... 56
 integrados..... 69
 limpar..... 75
 operacionais 69
 organizacionais 116
 preparar 161
 recolha 181
 repositórios 53
data marts 14, 69
data mining ..ix, 5, 8, 9, 10, 73, 74, 75, 77, 78, 79, 80, 82, 84, 95, 96, 97, 100, 101, 103, 104, 105, 106, 132, 180, 188
 algoritmos 73, 107
 aplicações..... 78
 estratégia 107
 ferramentas 4, 15, 44, 77, 80, 87, 88, 90, 110, 111, 114, 121, 122, 123, 131, 133, 134, 138, 142, 155, 161
 futuro 81
 modelos para representação do conhecimento
 obtido..... 82
 perito..... 96
 problemas não resolvidos..... 80
 processo 81
 utilizadores..... 75
 data warehouse..60, 62, 63, 64, 65, 67, 68, 69, 70, 75, 76, 77, 81, 100, 106, 107, 121, 180, 185
 DCBD Ver base de dados:descoberta do conhecimento
 dedução 8, 118
 descrição 45, 87
 elementar 87

dimensão temporal..... 109
 discretizar um atributo contínuo..... 75
 DM..... Ver *data mining*
 domínio
 da aplicação 117
 do problema 74
drill-down 107
dirty data..... 71

E

E..... Ver expressão
 EBNF ix, x, 113, 114
 EDI..... ix, x, 51
electronic data interchange Ver EDI
e-mail 14, 26
 empresa
 função 17
 bases de dados 41, 44, 158
 conhecimento dos clientes leais 36
 cultura..... 26
 gestor 156
 episódio 85
 espaço de pesquisa 73, 107
 especificação formal 116
 estatística..... 71
 estratégia 20
 estudos de casos 178
 evento 24, 31, 49, 85, 86, 157
 sequência 85
 evolução temporal..... 8, 46, 47, 63, 120, 181, 183
 expressão..... x, 83
extended Backus-Naur form..... Ver EBNF

F

factos..... 44, 69
 análise 45
 registo 48, 53
 FAIS..... 78
Falcon Credit-Risk Assessment..... 78
 ferramenta de aprendizagem 107
 folha 83, 89
 forma normal..... 58
 3ª 58
 formalização..... 132
Foxpro 161, 203
 função
 canónica..... 133, 143
 confronto 125, 133
interface..... 185
 funcionalidade..... 137, 182

G

GCCC Ver gestor de conhecimento:do comportamento dos clientes
generalised rule induction Ver GRI
 geração pós-relacional 58
 gestão

do conhecimento do comportamento dos clientes 3,14, 108,184
do relacionamento com os clientes 3, 180, 181
estratégias 180
gestor de conhecimento do comportamento dos
clientesix, x, 98, 104, 107, 109, 119, 120
construção 104
tarefas 132
GRI ix, x, 139, 225

H

hábitos
compra 33
consumo 30
utilização de serviços 33
HCI x
histórico das compras 21
human computer interface Ver HCI

I

imagens de satélites 50
inconsistência 56
incrementabilidade 112
information warehouse 62
inteligência artificial 71, 180
interface 6, 54, 62, 105, 112, 127, 129, 135
Internet 14, 26, 30, 39, 51, 89

L

lex 111, 122, 203
linguagem 8
hospedeira 123
natural 112
programação 117
lógica relacional 56

M

marketing 152, 180
campanhas .8, 23, 28, 31, 32, 33, 95, 97, 103, 139,
144, 147, 149, 153, 154, 156, 161, 178, 182,
183, 186
eficácia 11, 26, 103
futuras 154
implementar políticas 32
mecanismos de inferência .42, 44, 46, 48, 60, 70, 117
medida de semelhança 185
mercados
técnicas 78
método
de transformação ou redução de dimensionalidade
..... 75
tradicional de análise de dados 70
modelo
conceptual dos dados 54
hierárquico 55
lógico/dedutivo 55

orientado aos objectos 55
relacional 55, 58
extensões 55
modularidade 112
momento 142, 144, 147, 149, 163, 166, 169
motor de inferência 117

N

negócio
perito 6, 102, 104, 115
neurónio 82
nível
confiança xii, 94, 105, 114, 115, 151
suporte xii, 94, 105, 114, 115
normalização 58

O

OLAP ix, x, 59, 61
on-line analytical processing Ver OLAP
operador 20, 118, 126, 129
em 118
entao 118
evolucao 118
parametros 118
regra 118
se 118
outliers 75

P

padrão 1, 5, 8, 47, 71, 73, 80, 93, 94, 95, 101, 102,
105, 106, 107, 109, 115, 124, 125, 128, 130, 151,
182
cenário do tratamento 96
comportamento 8, 10, 14, 18, 28, 38, 40, 47, 67,
79, 104, 109, 121, 122, 124, 174, 181, 183
evolução temporal 94
compra 38, 47, 181
formato 121
registo 14
página Web 53, 95
redefinição 35
parametrização 158
Pascal 8, 117
Pentium 8, 111, 203
perfil 46, 144, 149, 151, 154
perito do domínio 127
ponto de contacto 19, 26
processo
aprendizagem 71
tomada de decisão 62, 69
produto
código 160
programa lógico 116
programação em lógica 116, 117, 180
Prolog 8, 111, 116, 117, 182
Win LPA 8, 203

protótipo....2, 8, 12, 14, 108, 111, 116, 117, 136, 138,
154, 175, 178, 182, 183, 184, 205
publicidade de produtos e serviços..... 35

R

rede neuronal..... 71, 82
redundância..... 54, 56, 58
registo56
regra 83
 actualização 107
 antecedente 88
 canónica 122, 132, 143, 150, 165
 canónica estendida 132, 134, 144, 146, 151, 165
 comportamento 5, 139, 153, 182
 consequente 83, 88
 corpo 83
 de associação 84, 88
 de classificação 84, 86
 descriptor 83
 eliminação 127
 evolução 106, 127
 grau de incerteza 114
 inferência 117
 interpretabilidade 84
 memorização 107
 opção negativa 152
 produção 8, 111, 112
 produção estendida 113
 semelhança 185
 sequencial 84, 86
regras semelhantes .124, 126, 131, 154, 162, 166, 174
regularidades 73
relação56
 causa-efeito 45
 igualdade 115
 inclusão 116
repositório 6, 62, 67, 70
 conhecimento 101
 dinâmico 101
roll-up 107

S

See5..... 79
segmentos de mercado 34
SGBD..... Ver bases de dados:sistema de gestão
sistema
 aprendizagem incremental 107

computacional..... 136
conhecimento.5, 6, 7, 9, 12, 14, 99, 111, 113, 120,
134, 181, 184
 arquitectura..... 6
 fechado 183
 identificação electrónica 50
 inferência 7, 186
 informação 71
 operacional 62
 recomendações..... 186
 viabilidade 136, 137, 178
SKICAT 79
SQL ix, x, 106, 161, 185
structured query language Ver *SQL*

T

tabela.....55, 56, 138, 158, 161
tarefas complementares 106, 185
tendências 30, 39, 45, 46, 73, 81
teoria da normalização 56
Time Inc 79
transações27, 34, 42, 51, 58, 70, 78, 85
transponder..... 50, 51
trigger 25

U

utilizador .6, 24, 36, 52, 53, 54, 59, 64, 67, 69, 75, 80,
84, 90, 95, 102, 104, 112, 120

V

viabilidade da solução..... 108

W

Windows NT..... 8, 111, 203

Z

ZACCARix, x, 5, 8, 9, 10, 15, 99, 100, 108, 109, 110,
115, 120, 123, 133, 136, 137, 139, 143, 144, 154,
155, 162, 178, 181, 182, 183, 205
 âmbito de aplicação 110
 viabilidade do protótipo 15
zelar pela aquisição do conhecimento dos clientes,
sua actualização e registo..... Ver ZACCAR