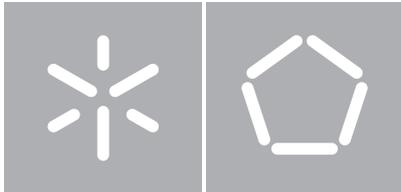


**Universidade do Minho**  
Escola de Engenharia



**Universidade do Minho**

Escola de Engenharia

Dissertação de Mestrado

Universidade do Minho  
Escola de Engenharia  
Departamento de Informática

Uma Ontologia para Expressão de Preferências de Utilizadores  
em Sistemas OLAP

**Eduardo Costa**

Dissertação de Mestrado

2012



# Uma Ontologia para Expressão de Preferências de Utilizadores em Sistemas OLAP

**Eduardo Costa**

Dissertação apresentada à Universidade do Minho para obtenção do grau de Mestre em Informática,  
na especialidade de Sistemas de Suporte à Decisão, elaborada sob orientação do Professor Doutor  
Orlando Manuel de Oliveira Belo.

2012



---

*Dedico esta dissertação aos meus pais, Eduardo Costa e Maria Araújo.*

---

---

# Agradecimentos

A dissertação que agora se apresenta resultou de um longo período de trabalho. Neste sentido, os méritos que ela possa ter, devem-se aos contributos das pessoas que durante, a sua elaboração, me proporcionaram testemunhos de vários géneros. Foram eles que a tornaram possível, expressando por isso a todos a minha gratidão.

Gostaria de destacar o papel desempenhado neste trajeto pelo meu orientador, o Professor Doutor Orlando Belo, que conquistou desde cedo a minha admiração pela forma entusiasta como lecciona as suas aulas, da forma como consegue facilmente chegar até ao aluno e principalmente pela excelente pessoa que é. Este foi um longo percurso, e nos momentos mais complicados estou-lhe mais reconhecido ainda, não só pelas dúvidas que me tirou, mas também pelo alento e força que me conseguiu transmitir.

O meu agradecimento vai, também para o conjunto de docentes do Mestrado de Informática da Universidade do Minho, o meu obrigado pelos saberes que me foram transmitidos.

Agradeço aos meus pais (José Eduardo Cunha da Costa e Maria Antónia da Silva Araújo), por terem sempre acreditado no seu filho, e por me ter incentivado e apoiado esta etapa da minha vida.

À minha namorada (Patrícia Alves), a pessoa que mais me incentivou durante este percurso, pelos sacrifícios a que esta dissertação obrigou. A sua companhia foi indispensável para conseguir chegar até este momento.

---

Por fim, mas não menos importante aos meus colegas de mestrado, principalmente ao Fernando Veloso que me acompanhou durante um longo período de tempo durante esta etapa.

---

## Resumo

### Uma Ontologia para a Expressão de Preferências de Utilizadores em Sistemas OLAP

Na maioria dos casos em que se submete uma query a um sistema, recebemos uma “enxurrada” de dados como resposta, que em grande parte não nos interessa. Na realidade, ao se obter um qualquer *record set*, fazemos frequentemente uma forte triagem para que fiquemos apenas com aquilo que realmente nos interessa no âmbito do problema que temos em mãos. De facto, era muito útil que, de algum modo, fosse possível expressar as nossas preferências em termos daquilo que queremos obter aquando do lançamento de uma dada query. Diversas pesquisas têm sido feitas com o objectivo de conseguir expressar as preferências dos utilizadores, no sentido de apresentar resultados mais satisfatórios nas suas consultas, mais próximos daquilo que o utilizador pretende. As preferências são utilizadas como filtragem nas consultas efectuadas, com o objetivo de reduzir o volume de dados que potencialmente poderá ser apresentado ao utilizador. Estas ajudam, também, na criação de políticas com base nos perfis dos utilizadores. Assim, neste trabalho de dissertação é proposto estudar a possibilidade de melhorar a interação dos utilizadores com os sistemas OLAP, recorrendo para isso à Web Semântica, nomeadamente à utilização de um modelo ontológico e a técnicas mais avançadas de expressão de preferências.

---

---

---

# **Abstract**

## **An Ontology for Expressing User Preferences on OLAP Systems**

In most cases in which a query is submitted to a system, the response is a “flood” of data which largely does not have any interest to us. Actually, to obtain any record set, we often do a strong triage to obtain what really interests us in the scope of the problem that we currently have. In fact, it would be very helpful that, somehow, we had the possibility to express our preferences in terms of what we want to get at the launch of a given query. Several researches have been done with the purpose of getting the user preferences expressed, in order to provide a more satisfactory result of the queries, closer to what the user pretends. User preferences are used as a filter in the result set, with the aim of reducing the data volume that could, potentially, be presented to the user. They help, also, in creating policies based on user profiles. Thus, in this dissertation work, it is proposed to study the possibility of improving the interaction of users with OLAP systems, resorting to Semantic Web, in particular the use of an ontological model and the most advanced techniques of expression of preferences.

---

---

---

# Índice

<b>1 Introdução .....</b>	<b>1</b>
1.1 Enquadramento do Trabalho .....	1
1.2 Motivação e Objectivos .....	3
1.3 A Estrutura da Dissertação .....	4
<b>2 Expressão de Preferências .....</b>	<b>7</b>
2.1 Personalização.....	8
2.1.1 Sistemas de Recomendação.....	9
2.1.2 Perfis de Utilização.....	12
2.1.3 Motores de Pesquisa .....	12
2.2 Preferências em Sistemas de Base de Dados.....	13
2.2.1 A Abordagem Quantitativa.....	15
2.2.2 Abordagem Qualitativa .....	17
<b>3 Preferências em Sistemas OLAP .....</b>	<b>23</b>
3.1 Preferências OLAP .....	23
3.1.1 A Abordagem Quantitativa.....	25
3.1.2 A Abordagem Qualitativa .....	27
3.2 Preferências com base em Ontologias .....	29
3.3 Sumário .....	33
<b>4 Ontologias .....</b>	<b>35</b>
4.1 Definição e Fundamentação.....	37
4.2 Razões para desenvolver ontologias.....	38
4.3 Metodologias para o desenvolvimento de ontologias.....	40
4.4 Linguagens de ontologias .....	43

---

4.4.1	Linguagens clássicas para a especificação de ontologias .....	43
4.4.2	Linguagens de especificação de ontologias com base na Web .....	44
4.5	Sumário .....	46
<b>5</b>	<b>Uma Ontologia para Preferências .....</b>	<b>47</b>
5.1	Definição do Domínio e do Âmbito .....	48
5.2	O Processo de Desenvolvimento .....	49
5.2.1	Determinar o âmbito e o domínio da ontologia .....	50
5.2.2	Considerar o uso de ontologias existentes.....	50
5.2.3	Enumerar os termos mais importantes.....	51
5.2.4	Determinar as classes e as hierarquias de classes.....	52
5.2.5	Definir propriedades.....	52
5.2.6	Definir restrições aos atributos.....	53
5.2.7	Criar instâncias .....	53
5.3	O Modelo Desenvolvido .....	53
5.3.1	Caracterização de um Perfil de Utilizador .....	54
5.3.2	Um Esquema Multidimensional.....	56
5.3.3	Reflexo das Preferências OLAP.....	59
5.4	Sumário .....	61
<b>6</b>	<b>O Caso de Estudo .....</b>	<b>63</b>
6.1	Apresentação e Caracterização do Problema .....	63
6.2	Motivação .....	64
6.3	Aplicação do Modelo .....	64
6.3.1	Perfil do Utilizador.....	65
6.3.2	O Esquema Multidimensional .....	67
6.3.3	Consultas e Preferências de Utilizador.....	68
6.4	Sumário .....	71
<b>7</b>	<b>Conclusões e Trabalho Futuro.....</b>	<b>73</b>
7.1	Um Comentário Final.....	73
7.2	O Modelo de preferências.....	74
7.3	Trabalho futuro .....	75
<b>8</b>	<b>Bibliografia.....</b>	<b>77</b>
<b>9</b>	<b>Anexo A .....</b>	<b>85</b>

---

---

## Índice de Figuras

Figura 1 – Diferenças entre pesquisas tradicionais e pesquisa personalizadas. ....	16
Figura 2 – Personalização do perfil de utilizador .....	16
Figura 3 – Introdução das preferências em sistemas de base de dados – figura adaptada de (Kießling, et al., 2001) .....	18
Figura 4 – Cubo de preferências apresentado em (Stefanidis, et al., 2005).....	26
Figura 5 – Detalhe do modelo de preferências apresentado em (Stefanidis, et al., 2005).....	27
Figura 6 – Modelo ontológico de uma preferência base retirado de (Toninelli, et al., 2008).....	31
Figura 7 – Modelo de preferências retirado de (Tapucu, et al., 2009).....	32
Figura 8 – Web Semântica .....	35
Figura 9 – Diagrama da Web Atual .....	36
Figura 10 – Diagrama da Web Semântica.....	37
Figura 11 - Estrutura da Web Semântica – adaptado de (W3C, 2007) .....	44
Figura 12 – O ambiente do Protégé para o desenvolvimento de uma ontologia.....	49
Figura 13 – Modelo de ontologia para o perfil de utilizador .....	54
Figura 14 – Exemplo de um perfil de utilizador .....	56
Figura 15 – O modelo de ontologia para o esquema multidimensional.....	57
Figura 16 – Exemplo da especificação de um esquema multidimensional.....	58
Figura 17 – O modelo de ontologia para preferências OLAP.....	59
Figura 18 – Exemplo de uma preferência .....	60
Figura 19 – Caracterização do perfil do 'Miguel Lopes'.....	64
Figura 20 – Caracterização do perfil do 'João Carvalho' .....	65
Figura 21 - Caracterização do perfil do 'Pedro Sousa' .....	66
Figura 22 – Caracterização do perfil da 'Raquel Cruz' .....	67
Figura 23 – O esquema multidimensional utilizado.....	68
Figura 24 – Modelo de ontologia para uma organização .....	70

---

Figura 25 – Caracterização dos agentes de decisão na organização “EJAC” .....	70
Figura 26 – O esquema multidimensional da “FoodMart” .....	85

---

## Índice de Tabelas

Tabela 1 – Linguagem SPARQL e Linguagem SQL.....	30
Tabela 2 – Estudo comparativo dos trabalhos realizados sobre preferências .....	32
Tabela 3 – Classes do modelo do perfil de utilizador .....	55
Tabela 4 – Classes do modelo do esquema multidimensional.....	57
Tabela 5 – Classes do modelo de preferências.....	59

---

---

---

## Siglas e Acrónimos

OLAP	-	Online Analytical Processing
DW	-	Data Warehouse
EM	-	Enterprise Model
TOVE	-	Toronto Virtual Enterprise
OCML	-	Operational Conceptual Modeling Language
W3C	-	World Wide Web Consortium
URI	-	Uniform Resource Identifier
XML	-	Extensible Markup Language
DL	-	Descriptive Logic
SPARQL	-	Simple Protocol and RDF Query Language
OWL	-	Web Ontology Language
RDF	-	Resource Description Framework
RDFS	-	Resource Description Framework Schema
XML	-	Extensible Markup Language
OIL	-	Ontology Interchange Language
DAML	-	The DARPA Agent Markup Language
GO	-	Gene Ontology
SQL	-	Structured Query Language
AIDAS	-	Adaptable Intelligent Discovery of context-Aware Services
SGDB	-	Sistema de Gestão de Base de Dados

---

---

# Capítulo 1

## Introdução

### 1.1 Enquadramento do Trabalho

Hoje uma organização não pode apenas olhar para si, deve ter sempre a noção do seu posicionamento relativamente à concorrência de forma a conseguir ser proactiva. Os mercados estão cada vez mais dinâmicos, exigindo frequentemente a integração de informação externa aos seus sistemas. Como sabemos, um dos factores principais que contribuiu para o “boom” dos sistemas de informação foi a Internet, onde a quantidade de páginas existentes atualmente é muito elevada. Neste momento, a Internet encontra-se com cerca de 125 milhões de domínios ativos (Kunder, 2007).

Numa fase inicial as organizações apenas utilizavam a Internet para publicitar a sua marca, os seus produtos e a suas localizações. Rapidamente se constatou o potencial que a Internet tinha e, de forma gradual, começaram-se a promover e a produzir aplicações de comércio electrónico (*e-commerce* e *m-commerce*). Organizações como Amazon<sup>1</sup>, Facebook<sup>2</sup> e eBay<sup>3</sup>, são apenas alguns exemplos onde o volume de dados existente nos seus sistemas de informação é bastante grande. Quando o utilizador (um agente de decisão) está a trabalhar neste tipo de sistemas de informação, vários problemas se colocam a este nível (Agrawal e Wimmers, 2000). A possibilidade de o utilizador efetuar uma consulta e não conseguir obter o resultado esperado é elevada. A forma

---

<sup>1</sup> <http://www.amazon.com/>

<sup>2</sup> <http://www.facebook.com/>

<sup>3</sup> <http://www.ebay.com/>

com que estes resultados surgem, sem qualquer tipo de organização e com um conteúdo de baixa importância, levou à necessidade de se conseguir formular consultas mais complexas, capazes de possibilitar a inclusão das próprias preferências dos utilizadores (Xin e Han 2008).

A ubiquidade dos equipamentos é também um factor que requer atenção especial na forma como a informação é disponibilizada em dispositivos móveis, uma vez que os equipamentos não possuem todos as mesmas capacidades em termos de hardware, tanto por limitações de processamento como de visualização dos dados. Os telemóveis, os *smartphones* e os *tablet's* são equipamentos cada vez mais utilizados como ferramentas de trabalho. As organizações quando desenvolvem um serviço para a Internet já não pensam apenas no seu acesso a partir de um qualquer *browser* num *desktop*, mas sim em tentar que este serviço fique disponível no maior número de dispositivos possível, quer seja na customização da página quer seja através de aplicações nativas.

Muitos dos utilizadores de sistemas gostariam de encontrar o melhor equilíbrio entre os seus desejos e a realidade. Contudo, a modelação das suas preferências não é algo de fácil concretização, uma vez que as preferências humanas são complexas, muito diversas, bastante heterogéneas e até mesmo, por vezes, contraditórias. Para além disso, elas são difíceis de avaliar pois o contexto em que estas foram expressas também pode variar, assim como o seu domínio de aplicação. A captação e a exploração das preferências de um utilizador têm sido propostas como uma solução para este problema em muitos domínios, em que, numa primeira fase, os trabalhos tiveram como base os processos de interação homem-máquina, os modelos de utilizador e a inteligência artificial (Gaasterland e Lobo, 1994). Posteriormente, seguiram-se alguns trabalhos em sistemas de base de dados (Agrawal e Wimmers, 2000; Kießling et al., 2001; Kießling e Köstler, 2002; Kießling, 2002; Chomicki, 2003; Koutrika e Ioannidis, 2004; Koutrika 2006), em sistemas OLAP (Stefanidis et al., 2005; Golfarelli e Rizzi 2009) e em Web Semântica (Siberski et al., 2006; Gursk et al., 2008; Toninelli et al., 2008; Tapucu et al., 2009). Embora as preferências em algumas dessas abordagens sejam definidas fazendo uso de uma ontologia, na maioria dos trabalhos citados anteriormente a noção de preferência foi integrada em domínios de aplicação bastante concretos. Vemos assim que, as preferências não podem ser partilhadas, porque são definidas e atualizadas para cada aplicação e domínios de aplicação distintos.

## **1.2 Motivação e Objectivos**

A Internet contribuiu para um crescimento exponencial dos dados e tornou-se numa ferramenta de comunicação e numa forma de comunicação indispensável. Contudo, a gestão da informação que nela é depositada é, claramente, limitada, uma vez que esta se encontra frequentemente em sistemas fechados ou em estruturas específicas que apenas algumas aplicações internas conseguem interpretar. Como consequência, a reutilização do conhecimento e uma possível extensão às estruturas torna-se praticamente impossível. O esforço coletivo entre várias organizações é apontado como um dos factores que mais contribui para o seu sucesso.

A Web Semântica surgiu em 2001 com o objectivo de conseguir melhorar a infraestrutura da Web e disponibilizando novos significados semânticos e novas interligações entre os dados. Em princípio, isto permite obter uma maior flexibilidade na utilização dos dados, possibilita a reutilização da informação e "abre" o conhecimento incorporado nos sistemas, permitindo aos humanos e aos computadores interagir e comunicar entre si. Porém, alguns desafios são colocados a este nível, como por exemplo aceder e manipular informação interligada na Internet ou conseguir encontrar informação relevante em fontes distribuídas.

Vários países tiveram que se adaptar ao processo de globalização, não existindo um consenso fechado quanto ao seu significado, nem à data em que terá ocorrido. De forma semelhante, existe a necessidade de uma "globalização" ao nível da própria Internet, quanto ao conhecimento que pode ser extraído, partilhado e reutilizado. Hoje em dia, as organizações estão espalhadas por diversas partes do mundo, trabalhando sempre que possível de forma colaborativa. Por sua vez, a informação encontra-se descentralizada havendo a necessidade de desenvolver modelos para que as várias fontes de informação consigam comunicar entre si. O desenvolvimento de modelos capazes de permitir e promover a partilha de conhecimento (salientando neste caso as preferências do utilizador) é uma mais valia para os sistemas de base de dados e, principalmente, para os sistemas OLAP devido à complexidade e volume dos dados que se encontram armazenados nos seus repositórios. Assim, com a realização dos trabalhos desta dissertação, pretendemos realizar uma:

- Análise às metodologias propostas até ao momento para a expressão de preferências.
- Análise ao conceito de ontologia, suas metodologias e linguagens de desenvolvimento.
- Definição de um modelo de ontologia formal, modular e extensível que:

- permita responder às especificidades de um utilizador OLAP.
- suportar o esquema de uma base de dados multidimensional.
- exprima capazmente as preferências de um utilizador de acordo com o seu contexto aplicacional.

De acordo com os vários objetivos apresentados, nesta dissertação propomos uma ontologia para representar as preferências dos utilizadores de sistemas OLAP. Além disso, pretendeu-se, também, efetuar um estudo detalhado relativamente à expressão de preferências e às suas metodologias e técnicas de especificação e desenvolvimento, de forma a alcançar um modelo para uma ontologia de preferências partilhável, formal e genérico.

### 1.3 A Estrutura da Dissertação

Para além do presente capítulo, esta dissertação está organizada em outros 5 capítulos, nomeadamente:

- **Capítulo 2** – *Expressão de Preferências* – É apresentado o conceito de preferências, com aplicação ao nível dos sistemas de base de dados tradicionais, enunciando os vários trabalhos efectuados. É, ainda, efectuada uma breve análise comparativa sobre as várias abordagens apresentadas.
- **Capítulo 3** – *Preferências em Sistemas OLAP* – Tal como no capítulo anterior, é apresentado o conceito de preferências, desta vez aplicado ao nível dos sistemas OLAP.
- **Capítulo 4** – *Ontologias* – Aqui é apresentado o conceito de ontologia, bem como as metodologias e as linguagens utilizadas no seu desenvolvimento. Além disso, apresenta-se a motivação para a criação e utilização de ontologias em cenários de aplicação real.
- **Capítulo 5** – *Uma Ontologia para Preferências* – Neste capítulo apresenta-se o modelo da ontologia proposto neste trabalho de dissertação e uma descrição detalhada quanto ao seu desenvolvimento. Posteriormente, efetua-se uma análise explicativa de cada um dos módulos - perfil de utilizador, esquema da base de dados multidimensional, preferências OLAP e contexto – relacionados com o modelo proposto.
- **Capítulo 6** – *O Caso de Estudo* – Tal como a designação do próprio capítulo indica, é apresentado o caso de estudo utilizado para ilustrar a forma como a ontologia proposta

pode contribuir para a melhoria da partilha de informação e para permitir a um utilizador obter melhores resultados durante uma sessão de exploração OLAP.

- **Capítulo 7 – Conclusões e Trabalho Futuro** – Por fim, são apresentadas as conclusões sobre o trabalho realizado nesta dissertação e indicadas algumas possíveis linhas de orientação para uma continuação do trabalho realizado até ao momento.



## Capítulo 2

### Expressão de Preferências

Muito trabalho tem sido feito ao longo das últimas duas décadas, com o objectivo de encontrar métodos capazes de expressar as preferências de cada utilizador de um sistema. De salientar, os trabalhos realizados no domínio da interação homem-máquina, dos modelos de utilizador e da inteligência artificial (Gaasterland e Lobo, 1994). O que se tem vindo a fazer é trabalhar em um conjunto de metodologias capazes de expressar as preferências e as necessidades do utilizador, com o objectivo de melhorar a sua interação com as aplicações. Estas pesquisas variam muito em termos das abordagens adotadas e na forma como a informação do utilizador é extraída. Tudo isto é, obviamente, dependente da área em que se está a trabalhar (motores de pesquisa, personalização de conteúdos, OLAP, etc.).

Numa primeira fase, os estudos da aplicação de preferências de utilizador foram direccionados para os sistemas de base de dados e, apenas, mais recentemente para os sistemas OLAP (a área de trabalho desta dissertação). Como resultado dos estudos efectuados, atualmente já existe um grande conjunto de abordagens para o tratamento de preferências no domínio dos sistemas de base de dados. As várias abordagens existentes, tanto para os sistemas de base de dados como para os sistemas OLAP, diferem em vários aspectos. De referir (Golfarelli et al. 2010):

- *Esforço de expressão.* Algumas abordagens exigem que o utilizador especifique manualmente os critérios de preferência para cada consulta, enquanto que em outros os critérios de preferência são estabelecidos com base no contexto e no perfil do utilizador.
- *Prescritivo.* Algumas abordagens utilizam critérios de preferência fortes (estabelecidos na

condição WHERE e que são adicionados a uma consulta. Outros, porém, fazem uso de critérios leves, isto é, utilizam novas cláusulas que introduzem a noção de preferências às *queries*.

- *Espírito de iniciativa*. Existem abordagens nas quais as consultas são efectuadas com base no histórico de navegação e do contexto e, posteriormente, devolvem o resultado ao utilizador. Assim a consulta efectuada é enriquecida com informação que ajudará a obter uma melhor resposta ao utilizador.
- *Expressividade*. Critérios de preferência têm expressividades diferentes e podem ser combinados de formas diferentes.

## 2.1 Personalização

A personalização é a adaptação e a disponibilidade dos serviços e informação de acordo com cada utilizador ou grupo de utilizadores. Isto implica que se faça uma optimização permanente às informações apresentadas ao utilizador. O utilizador não só beneficia obtendo informação que lhe é mais relevante, como também existe um tipo de "sentimento de familiarização" do utilizador com o próprio sistema.

Nos dias de hoje, o acesso à Internet está disponível através de vários equipamentos e plataformas. Os telemóveis e os *smartphones* apresentam percentagens de utilização quase ao nível dos computadores (tipo desktop e portátil). O consumo de conteúdos na Internet através da televisão também é cada vez mais um realidade. As características físicas de todos estes equipamentos obrigam a uma alteração da forma como os conteúdos são apresentados, como também a quantidade da informação a ser visualizada sofre alterações. Esta mudança de paradigma obriga a um esforço adicional por parte de quem está a desenvolver sistemas para a Web, para que se possa consumir os seus diversos tipos de conteúdos nas várias plataformas e dispositivos disponíveis aos utilizadores comuns.

A personalização tornou-se numa ferramenta fundamental para a resolução dos problemas indicados. No entanto, existem ainda algumas dificuldades neste domínio, que têm a ver, essencialmente, com:

- *A violação de privacidade*. É importante que o utilizador não tenha a sensação que a sua

- privacidade é violada de alguma forma.
- *A personalização errada.* Evitar que exista uma má experiência com um processo de personalização, já que isso pode fazer com que o utilizador deixe de utilizar o serviço.
  - *O armazenamento dos perfis de utilizadores.* Uma vez que o perfil dos utilizadores está armazenado do lado de quem está a oferecer o serviço, muitas das vezes esta informação é utilizada de forma indevida sem o utilizador ter conhecimento em situações como:
    - *O spam.* Utilização do email do utilizador para envio de emails de publicidade, sem que este seja notificado.
    - *A venda de informação a terceiros.* A venda de informações pessoais a terceiros é efectuada com bastante frequência, já que é uma forma fácil de ganhar dinheiro.
  - *As recomendações.* Os próprios proprietários dos produtos / serviços podem gerar várias recomendações positivas para si e efetuar recomendações negativas para os seus concorrentes.

As redes sociais, como o Facebook, possuem atualmente um valor de mercado enorme, devido em grande medida à importância que é atribuída à informação pessoal do utilizador nos dias de hoje, envolvendo inúmeros aspetos informativos como os seus interesses, atividades, contactos, preferências, e muitas coisas mais. Já em (Resnick, 1997) a posse da informação pessoal do utilizador era vista como uma mais valia para os sistemas, mas a ao mesmo tempo, também, como um problema. Isto porque, na eventualidade de a informação estar a ser vendida, dever-se-ia dar uma qualquer compensação monetária ao utilizador. Por um lado, o utilizador consegue obter um melhor serviço ao contribuir com informação para o sistema, mas, por outro, a organização sai a beneficiar pois não só consegue desenvolver um melhor produto/serviço como também pode obter lucros das vendas. Os sistemas que fazem uso da personalização para melhorar a experiência do utilizador têm, acima de tudo, que transmitir confiança na sua utilização e não fazer um uso abusivo da informação que tem em suas mãos.

### **2.1.1 Sistemas de Recomendação**

Os sistemas de recomendação (Resnick e Varian, 1997) têm sido uma área de investigação bastante popular, desde que se constatou que na Internet pessoas com diferentes origens seriam capazes de consultar a mesma informação. Muitas vezes somos deparados com várias alternativas em termos de escolha, mesmo em situações em que não temos qualquer experiência ou

conhecimento da matéria envolvida. Na vida quotidiana, contamos com recomendações de outras pessoas sobre as mais diversas situações, sejam elas realizadas por via oral, por jornais ou pela Internet, como é o caso, por exemplo, de uma recomendação de um filme ou de um restaurante. Os sistemas de recomendação ajudam nestes processos sociais, produzindo previsões, recomendações e opiniões que ajudam o utilizador a avaliar e a seleccionar objetos, produtos ou serviços. A ideia básica de um sistema de recomendação é a de que um utilizador selecione objetos, marque-os ou consulte-os, e, depois, o sistema identifica outros objetos semelhantes, a partir dos quais ele possa fazer recomendações (ou previsões) sobre aquilo que o utilizador possa gostar (Koutrika, 2006).

Estes sistemas variam muito quanto á forma como ajudam o utilizador e às tecnologias que utilizam para atingirem os seus objectivos. As duas abordagens mais populares são os sistemas de recomendação com base em conteúdos (Kalles et al., 2003) e a filtragem colaborativa (Herlocker et al., 2000; Hofmann 2003). Contudo, outras técnicas híbridas tem sido também propostas. Veja-se o caso de (Balabanovic e Sholam, 1997).

Os sistemas de recomendação com base em conteúdos são utilizados quando a recomendação tem como base o conteúdo de documentos de texto, tais como páginas Web, publicações ou notícias. É necessário existir alguma informação sobre as preferências do utilizador. Normalmente essa informação é obtida durante um qualquer processo de registo. As recomendações são realizadas tendo por base o conteúdo dos documentos, existindo uma filtragem consoante as preferências que o utilizador estiver a utilizar no sistema.

A filtragem colaborativa utiliza o conhecimento de preferências de grupos de utilizadores para fazer recomendações ou previsões sobre as preferências de outros utilizadores cujos perfis sejam semelhantes. Este sistema de recomendação tornou-se bastante popular pela sua simplicidade e pelos resultados satisfatórios em termos de desempenho, face ao volume de dados com que trabalha (Schickel-Zuber e Faltings, 2006). As principais abordagens de filtragem colaborativa são (Su e Khoshgoftaar 2009):

- *A abordagem com base em memória.* Esta abordagem é das mais populares em processos de filtragem colaborativa. A ideia por trás desta abordagem é simples: utilizar os dados que contêm a pontuação do utilizador para calcular a similaridade entre utilizadores e

produtos. As principais vantagens desta técnica inclui: ser de fácil implementação; a incorporação de informação nova é fácil e pode ser realizada de forma incremental; não é necessário considerar o conteúdo dos que estão a começar a ser recomendados. As principais desvantagens são: é uma abordagem que está dependente da pontuação dos utilizadores; a sua performance diminui quando a informação é escassa; não consegue recomendar para novos utilizadores e produtos; e é pouco escalável para grandes quantidades de dados.

- *A abordagem com base em modelos.* Aqui, os modelos são desenvolvidos utilizando algoritmos de mineração de dados e máquinas de conhecimento para encontrarem padrões com base em dados de treino. Existem vários algoritmos de filtragem colaborativa com base em modelos, que incluem redes bayesianas, modelos de agrupamento (*clustering*), processos de decisão de Markov com base em modelos, entre outros. Esta abordagem tem um objectivo mais holístico para descobrir factores escondidos nas pontuações dos utilizadores. As principais vantagens desta abordagem são: lida melhor com problemas como a escassez de informação, escalabilidade, entre outros; melhora o desempenho da previsão; e dá uma justificação intuitiva para as recomendações. As principais desvantagens são: o custo de construção do modelo é elevado; existe um compromisso entre performance de previsão e escalabilidade; e pode-se perder informação útil, devido á redução dos modelos.
- *Abordagem híbrida.* Esta abordagem pretende superar as limitações e desvantagens das abordagens com base em memória e com base em modelos. O custo da implementação desta abordagem é elevado e o seu nível de complexidade aumenta.

No entanto, alguns problemas foram identificados na filtragem colaborativa, como, por exemplo, ser pouco escalável, no início (da aplicação e de um novo utilizador) não existir informação suficiente para conseguir encontrar grupos com gostos semelhantes (*cold start*), e o problema de não conseguir recomendar absolutamente nada até obter a sua primeira pontuação (*first rater*) (Middleton et. al., 2002). A maior diferença entre a filtragem colaborativa e os sistemas de recomendação com base no conteúdo é que a filtragem colaborativa depende apenas dos dados com a pontuação dos utilizadores, enquanto que a recomendação de sistemas com base no conteúdo depende das características dos utilizadores e produtos para conseguir efetuar as suas previsões.

### 2.1.2 Perfis de Utilização

A construção do perfil de um utilizador é um elemento fundamental na personalização. É caracterizado, usualmente, por um conjunto de informação pessoal, constituída por uma descrição de capacidades físicas e intelectuais, preferências, atividades, conhecimentos, estilos de vida e interações com sistemas (Fröschl, 2005). Se não existisse informação sobre os utilizadores, os sistemas tinham o mesmo comportamento e o mesmo aspecto para todos os utilizador. Desta forma é possível fazer a personalização para cada utilizador, para um determinado produto ou serviço. As técnicas mais comuns na construção de um perfil de utilizador são (Tapucu, et al. 2009):

- *Explícita*, na qual cada utilizador é questionado num formulário quando visita um site. Este método tem a vantagem de deixar o utilizador especificar os seus interesses diretamente.
- *Implícita*, em que os sistemas obtém as informações dos utilizadores com base nas suas ações de navegação. Este método é normalmente transparente para o utilizador.

A experiência mostra que são raros os utilizadores que definem o seu perfil de forma explícita, não só pelo tempo que leva a definir as propriedades, como também pelo facto de nem sempre o ambiente de personalização ser o mais amigável possível. Assim, os serviços tem que recorrer frequentemente a técnicas automáticas de recolha de informação com base na navegação e nas ações realizadas pelos utilizadores.

### 2.1.3 Motores de Pesquisa

Como já referido, a informação presente na Internet é grande e o seu crescimento é rápido. Os motores de pesquisa são uma ferramenta fundamental na aproximação da informação ao utilizador, recebendo atualmente milhares de consultas por dia. Os utilizadores tendem a efetuar consultas com textos curtos, recorrendo a palavras-chave e, na sua maioria, nunca vão para além da primeira página de resultados que é apresentada. Isto obriga a que os motores de pesquisa tenham algoritmos eficientes e apresentem resultados com um bom nível de exatidão.

O motor de pesquisa da Google<sup>4</sup> é atualmente o mais utilizado no mundo, em parte devido às suas

---

<sup>4</sup> <http://www.google.com>

características e à sua "grandiosidade". Alguns dos seus utilizadores até o designam como a própria Internet. O Google surgiu em 1996, num projeto de investigação de dois alunos de doutoramento da Universidade de Stanford, Lawrence Page e Brin Sergey. Porém, só em 2001 se tornou público. Sergey e Page estavam com um sentimento de frustração com a utilização do AltaVista<sup>5</sup>, o melhor motor de pesquisa da altura. Embora a navegação pela Internet se revelasse bastante razoável nesses tempos, a hierarquização dos resultados da pesquisa efetuada era má. O PageRank (Page, et al., 1999) foi o algoritmo criado pelos dois alunos que mudou praticamente tudo. Este consiste, basicamente, num método de atribuir uma pontuação às páginas Web, de forma objectiva e mecânica, medindo assim o interesse dos utilizadores e a atenção dedicada a cada uma delas. Este modelo tem como base o modelo de avaliação dos artigos científicos, que atua mais ou menos assim: quanto mais citações um artigo científico tiver, mais provável será que seja um bom artigo. Da mesma forma, as páginas que mais eram referenciadas na Internet eram apresentadas em primeiro lugar pelo Google. Este problema de hierarquização de resultados está presente em muitas aplicações. O Google é frequentemente referenciado pois no seu domínio de atuação consegue-o fazer muito bem. Por vezes, chegamos mesmo a perguntar "Como é que ele sabia isto?", já que ele desenvolve o processo de personalização, com base no perfil de um utilizador, sobre as pesquisas que lhe pedem. No entanto, nem todos os sites oferecem personalização ou outro tipo de customização. Isso torna as coisas um pouco mais difíceis. Ainda hoje, é difícil fornecer uma solução de personalização para cada utilizador de um sistema. Alguns sites optam por oferecer um conjunto genérico de funções personalizáveis de forma a conseguir satisfazer (em parte) os seus utilizadores.

## **2.2 Preferências em Sistemas de Base de Dados**

O primeiro trabalho sobre preferências em sistemas de base de dados foi realizado por Lacroix e Lavency (1987), trabalho ao qual na altura não foi dada a devida importância. Pouco tempo depois, em 1993, na Universidade de Augsburg, o tema desencadeou várias pesquisas à sua volta e em diversas áreas. Com uma visão de pesquisa a longo prazo, esta foi nomeada de "It's a Preference World" por Hafenrichter e Kießling (2005). Estes autores não estavam só interessados na aplicação das preferências dos utilizadores para melhorar o resultado das consultas efectuadas a sistemas de base de dados, mas também em estudar as preferências em áreas como a:

---

<sup>5</sup> <http://www.altavista.com/>

- Automatização do processo de vendas B2B (*Business to Business*) – *P-Bargainer, P-Agent*.
- Aplicação das preferências em servidores aplicativos J2EE – *P-Services*.
- Consulta de documentos – *P-News*.
- Linguagens de consulta (SQL, XML) – *Preference XPATH*.
- Performance e otimização de serviços – *P-Optimizer, SR-Combine*.

No que diz respeito ao processo de tratamento de preferências em sistemas de base dados, este costuma ser abordado de duas formas distintas: a abordagem quantitativa e a abordagem qualitativa. Estas duas abordagens divergem quanto à maneira como as preferências são expressas. A forma como o utilizador expressa as suas preferências é através do recurso de um formulário que é preenchido ou através do seu perfil de exploração. Os motores de pesquisa que estão implementados com SQL *standard* sofrem da falta de noção de preferência, o que leva o utilizador a agir como as base de dados SQL, através de restrições duras implementadas pela condição WHERE.

**Exemplo:** *O João deseja comprar um carro usado. Deve ser um BMW. Além disso, o carro deve ter cor cinza ou preta. Igualmente importante é o preço do carro, que deve rondar os 30,000.00€. Também não deve ter um consumo de combustível alto. No entanto, esta, não é uma característica tão importante.*

Querer um BMW é uma condição dura ou fixa. Porém, o João estabelece um condição leve "deve ser um BMW". Mas, poderia existir um carro, de uma outra marca, que fosse igualmente interessante. Da mesma forma, para as restantes preferências estabelecidas pelo João, deve existir um compromisso na resposta que é dada ao utilizador, para que, mesmo não conseguindo encontrar a resposta ideal, se conseguisse encontrar a melhor alternativa possível e, claro, que o utilizador fique satisfeito. As organizações têm como principal objectivo vender os seus produtos, mas associado a isso, querem também os seus clientes satisfeitos. Como tal, não encontrar um qualquer produto, que devido a critérios de filtragem não foi apresentado ao utilizador, tem um impacto bastante negativo na sua apreciação. Isto, com a devida adaptação, pode ocorrer num processo típico de consulta de dados.

### 2.2.1 A Abordagem Quantitativa

Na abordagem quantitativa são usadas funções de pontuação para atribuir um valor de preferência. Esta abordagem possui algumas limitações, principalmente quando o volume de dados é grande – a falta de expressividade da abordagem quantitativa está bem presente na teoria da utilidade<sup>6</sup> apresentada por Chomicki (2003). O exemplo que de seguida e apresenta ilustra a aplicação prática desta abordagem:

**Exemplo:** *Considerando uma lista de carros com o objectivo de identificar as preferências de um utilizador, este atribui um valor de pontuação a cada um dos carros, sendo depois seleccionados aqueles que tiveram uma nota mais elevada.*

Em (Agrawal e Wimmers, 2000) foi apresentado um modelo seguindo esta abordagem, que serviu de referência para investigações efectuadas posteriormente. Aí o autor propõe um modelo no qual a preferência do utilizador é expressa por um valor real, num intervalo de valores entre 0 e 1, em que 0 representa falta de interesse e 1 representa elevado interesse. Fundamentalmente, este modelo consiste em duas noções básicas:

- Uma função de preferência que especifica a preferência do utilizador.
- Uma combinação de funções, de forma a se conseguir definir as preferências quando existe mais do que um utilizador.

Uma outra perspectiva surgiu mais tarde nos trabalhos de Koutrika e Ioannidis (2004) e de Koutrika (2006), em que a expressão de preferências era baseada no perfil do utilizador. O modelo aí proposto assentava, sobretudo, em elementos atómicos de *queries* (condições de selecção e de junção). A Figura 1 ilustra a alteração introduzida quando é efectuada uma consulta seguindo o modelo apresentado. Do lado esquerdo dessa figura apresenta-se uma consulta tradicional e do lado direito uma consulta fazendo uso do perfil de utilizador.

---

<sup>6</sup> A teoria da utilidade consiste na representação das preferências relativas de um indivíduo, entre os elementos de um conjunto.

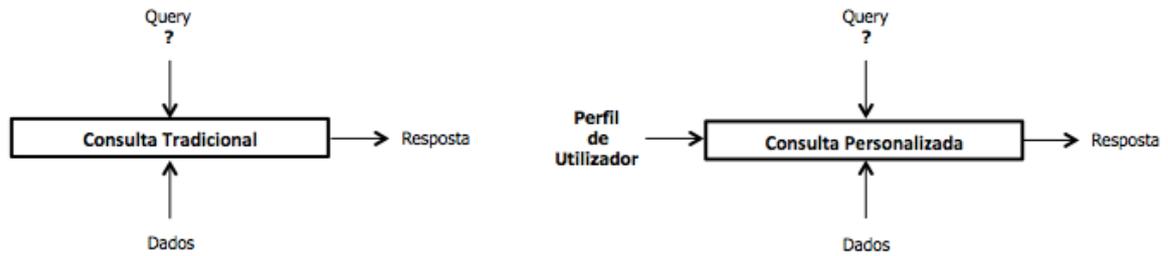


Figura 1 – Diferenças entre pesquisas tradicionais e pesquisa personalizadas.

O perfil do utilizador proposto, tinha comi base a construção de um grafo  $G<V,E>$ , no qual V indica os vértices que neste contexto representam as relações entre entidades, atributos e valores pelos quais o utilizador se interessa, e E são as arestas que neste contexto representam a seleção de atributos e junção entre entidades. As preferências para valores de atributos são guardados como seleções atômicas, enquanto que as preferências para as relações entre entidades são guardadas como condições de junção atômicas.

**Exemplo:** Considerando o esquema de base de dados: Filme (id, nome, duração, ano, did, gid); Diretor (id, nome); Gênero (id, nome); a Figura 2(b) mostra o gráfico de preferências de acordo com o perfil do utilizador Figura 2(a).

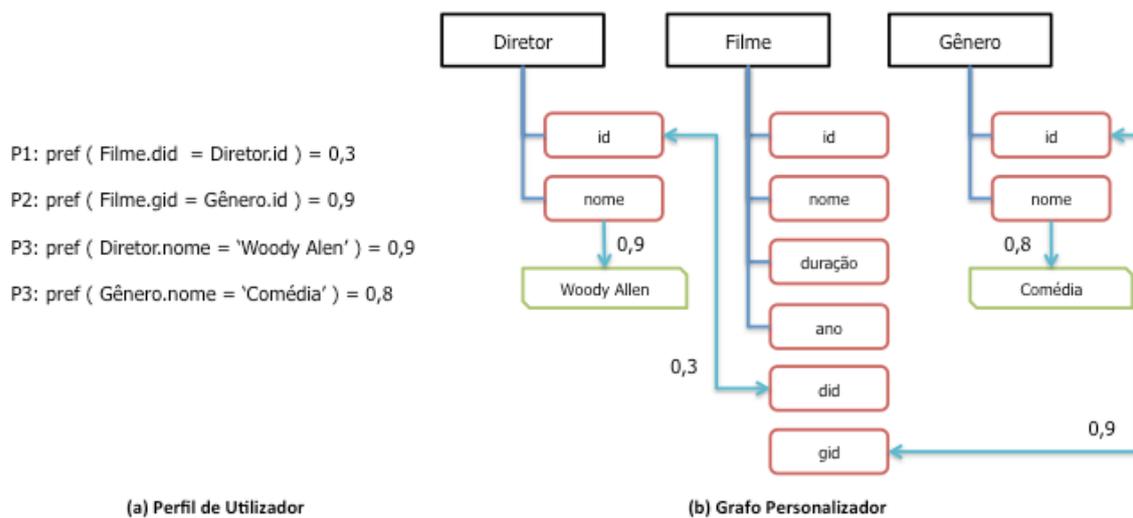


Figura 2 – Personalização do perfil de utilizador

Desta forma, o utilizador ao efetuar uma *query*, numa primeira fase, faz com que sejam consultadas as preferências do utilizador e executados algoritmos eficientes para fazer a computação sobre o grafo de personalização. A *query* é personalizada, executada e, posteriormente, e o seu resultado devolvido ao utilizador.

### 2.2.2 Abordagem Qualitativa

As preferências tiveram um papel bastante importante na última década, principalmente no ramo das ciências sociais. Contudo, só recentemente é que se conseguiu definir um modelo capaz de cobrir de forma eficiente a integração das preferências nos sistemas de base de dados. As pessoas expressam frequentemente as suas preferências de forma comparativa (ex. "Eu gosto mais de A do que de B"). Nem sempre é intuitivo o uso de termos numéricos, ou seja, atribuir uma dada pontuação na expressão da sua preferência. No entanto, isso não deixa de ser um factor bastante importante que é usado com bastante frequência, no qual o factor numérico tem que estar presente na definição de preferência. Esta abordagem tem como base a teoria da ordem<sup>7</sup> da matemática, mais concretamente, relações de ordem parcial.

Chomicki (2003) e Kießling (2002) tratam as preferências como ordens parciais estritas (cada preferência deve ser irreflexiva e transitiva, logo assimétrica). No entanto a abordagem efectuada por Chomicki define fórmulas mais genéricas. A prioridade das preferências também é uma questão abordada por ele, mas neste caso não faz uso das preferências Pareto (operador introduzido na expressão de preferências para definir prioridades entre elas). Ambos os autores sugerem novos operadores que, na sua essência, estendem a linguagem SQL standard, e cuja semântica é bastantes parecida. Segundo Kießling (2002) para chegar ao modelo capaz de expressar tanto preferências simples como complexas, tem que se ter as seguintes características:

- *Semântica intuitiva.* A forma como se expressa as preferências deve ser intuitiva e ter uma especificação declarativa.
- *Fundamento matemático.* Uma preferência deve ter uma base matemática, para definição da sua semântica.

---

<sup>7</sup> Teoria da ordem é um ramo da matemática que estuda vários tipos de relações binárias que capturam a noção intuitiva de um matemático de ordenação.

- *Um modelo de preferência construtivo e extensível.* As preferências complexas obrigam a um modelo que permita uma construção dinâmica, ou seja, o modelo tem que ser capaz de interpretar um conjunto de preferências.
- *O conflito de preferências não causa falhas no sistema.* Perante um modelo de construção de preferências dinâmico podem existir conflitos - duas pessoas ao expressar as suas preferências podem entrar em conflito. Daí a necessidade de o modelo contemplar este tipo de situações para não provocar falhas ao sistema.

Kießling (2002) apresenta um modelo que diz ser a chave para uma engenharia e álgebra de preferências. O modelo apresentado estende as abordagens existentes para rankings numéricos e não numéricos e tenta introduzir um novo conceito designado como engenharia de preferência. Dois tipos de preferências surgem neste modelo, as preferências base e as preferências complexas, onde a sua explicação é apresentada mais à frente.

Em 1997 apareceu uma ferramenta, designada por *Preference SQL (Standard SQL + Preference)*, que introduziu de uma camada intermédia entre as aplicações e os sistemas de base de dados, no qual as consultas que expressassem preferências dos utilizadores eram traduzidas em consultas SQL tradicionais (Kießling e Köstler 2002). Apenas mais tarde, em 1999, surgiu como um produto comercial. Da mesma forma que surgiu o *Preference SQL*, surgiu também o *Preference XPATH*, que seria aplicado em sistemas de base de dados baseados em XML (Kießling, et al., 2001). A Figura 3 ilustra a forma como foram introduzidas as preferências nos sistemas de base de dados.

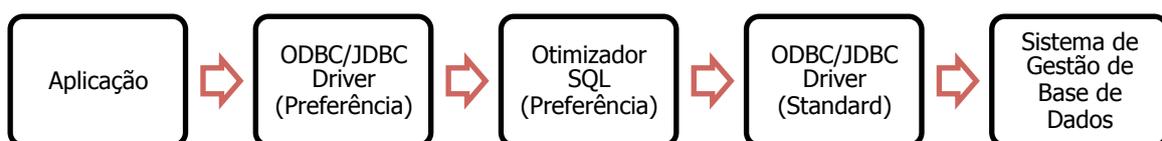


Figura 3 – Introdução das preferências em sistemas de base de dados – figura adaptada de (Kießling, et al., 2001)

Outras abordagens foram efectuadas, como aquela que definiu o operador SKYLINE (Börzsönyi, et al., 2001). Basicamente, esta abordagem estende a SQL com uma nova cláusula: SKYLINE OF (ver exemplo apresentado de seguida). Deste trabalho surgiram outros algoritmos mais eficientes,

capazes de suportarem esta forma de expressar preferências de utilização. Este novo operador teve como (grande) vantagem a não alteração do modelo subjacente à SQL até então estabelecido. No entanto, não conseguiu cobrir toda a definição de preferências de um utilizador.

**Exemplo:** *"Um utilizador procura um carro da marca BMW no qual consiga obter a melhor relação entre a potência e o consumo".*

```
SELECT * FROM Carros
WHERE marca = 'BMW'
SKYLINE OF potência MAX, consumo MIN;
```

Devido à importância do modelo de preferências desenvolvido por Kießling (2002), que serviu como base para muito dos trabalhos que se seguiram, nomeadamente para o modelo ontológico de preferências apresentado neste trabalho de dissertação, iremos aprofundá-lo e explicar como se expressam as preferências do utilizador. Salientar, novamente, que as preferências segundo

### Preferências Base

As preferências base consistem na representação de uma preferência.

– Para atributos categóricos:

- POS(A, POS-set): especifica que um determinado conjunto de valores deve ser preferido - e.g. POS(transmissão, automática).
- NEG(A, NEG-set): especifica que um determinado conjunto de valores deve ser evitado, se possível - e.g. NEG(marca, Ferrari, Porsche).
- POS/NEG(A, POS-set; NEG-set): expressa as preferências de forma combinada, especificando um conjunto de valores que deve ser preferido e, por sua vez, o conjunto de valores que deve ser evitado - e.g. POS/NEG(cor, amarelo; vermelho).
- POS/POS(A, POS1-set; POS2-set): especifica quais os valores de dois conjuntos que devem ser preferidos - e.g. POS/POS(categoria, *cabriolet*; *roadster*).
- EXP(A, E-graph): enumera explicitamente as relações "gosto mais do que" - e.g. EXP(cor, (verde, amarelo), (verde, vermelho), (amarelo, branco)).

– Para atributos numéricos:

- AROUND(A, z): especifica a preferência sobre os valores mais próximos de um

determinado valor -  $\text{AROUND}(\text{preço}, 40000)$ .

- $\text{BETWEEN}(A, [\text{min}, \text{max}])$ : especifica a sua preferência sobre valores que se devem posicionar num dado intervalo de valores -  $\text{BETWEEN}(\text{quilómetros}, [20000, 30000])$ .
- $\text{LOWEST}(A)$ : especifica a sua preferência sobre valores mais baixos -  $\text{LOWEST}(\text{consumo})$ .
- $\text{HIGHEST}(A)$ : especifica a sua preferência sobre valores mais altos -  $\text{HIGHEST}(\text{potencia})$ .
- $\text{SCORE}(A, f)$ ,  $f: \text{dom}(A) \rightarrow \mathbb{R}$ ,  $P$  é designada por preferência SCORE, se para  $x, y \in \text{dom}(A)$ :  $x < P y$  sse  $y < P x$ .

### Preferências Complexas

Uma preferência complexa consiste numa expressão de várias preferências base, utilizando um operador que define a importância e a sua ordem. Vejamos então os tipos de preferências de uma forma mais concreta. Dado  $P_1 = (A_1, < P_1)$  e  $P_2 = (A_2, < P_2)$ , uma:

- Preferência Pareto:  $P := P_1 \otimes P_2 \otimes \dots \otimes P_n$ , em que  $P$  é a combinação de  $n$  preferências sobre as quais este operador estabelece a mesma importância. Isto significa
- Preferência prioridade:  $P := P_1 \& P_2 \& \dots \& P_n$ , em que  $P$  é a combinação de  $n$  preferências sobre as quais este operador estabelece uma ordem de importância. Isto significa, para o caso da expressão apresentada que  $P_1$  é mais importante do que  $P_2$ , por sua vez  $P_2$  é mais importante do que  $P_3$ , onde  $P_n$  é a última preferência a ser avaliada.
- Preferência numérica:  $P := \text{rank}_f(P_1, P_2, \dots, P_n)$ , em que  $P$  combina preferências SCORE  $P_i$ , através de uma função  $F$  numérica de ranking.

Abaixo é ilustrado um exemplo que pretende demonstrar a aplicação dos conceitos inerentes ao modelo.

**Exemplo:** *"Sob nenhuma circunstância a Joana pode gastar mais do que 35,000€ na compra de um carro. Ela deseja um carro que deve ser um BMW ou um Mercedes, que deve ter por volta de 5 anos e uma cor que não pode ser vermelha. Todas as preferências são igualmente importantes para ela."*

$$P_{Joana} = POS(marca, \{ 'BMW', 'Mercedes' \}) \otimes AROUND(idade, 5) \otimes NEG(cor, 'Vermelho')$$

A consulta que expressa a preferência da Joana, que vai dar origem ao melhor conjunto de carros sobre o qual se encaixam as suas preferências, poderá ser:

```
SELECT c.preço, cat.marca, c.idade, cat.cor
FROM carros_usados c, categoria cat
WHERE c.ref = cat.ref AND c.preço <= 35000
PREFERRING c.marca IN ( 'BMW', 'Mercedes' ) AND
             c.idade AROUND 5 AND
             cat.cor NOT IN ( 'Vermelho' );
```

A forma como se consegue expressar a *query* é bastante intuitiva e segue o raciocínio lógico que as pessoas têm. A introdução do operador *Pareto* (ilustrado no exemplo que dá origem no SQL ao operador AND) permite expressar preferências como o mesmo nível de igualdade, assim como a ordenação da importância é conseguida através do operador CASCADE, que no exemplo presente não se aplicou, pois todas as preferências da Joana tem a mesma importância.



## Capítulo 3

### Preferências em Sistemas OLAP

#### 3.1 Preferências OLAP

As bases de dados multidimensionais são repositórios especializados que possuem grandes volumes de dados, sobre as quais são realizadas análises complexas para suporte a processos de tomada de decisão. Geralmente, a implementação das bases de dados multidimensionais é pensada e implementada sobre uma determinada área de negócio, sendo organizadas por temas de análise, refletidos em registos de dados designados por factos, e por eixos de análise, suportados por atributos especiais designados por dimensões. Por sua vez, as dimensões estão organizadas por hierarquias, que estabelecem o nível de granularidade com que se pode consultar a informação (com maior ou menor detalhe).

O termo *data warehouse* surgiu por volta de 1990 introduzido por Bill Inmon (CIF, 2007), que representa um repositório de dados especializado no armazenamento de informação de forma integrada, estruturada. É uma estrutura não volátil, que tem como base suportar os processos de tomada de decisão de uma organização (Gruber, 2009). Vulgarmente criados sobre os *data warehouses*, os sistemas OLAP (*On-Line Analytical Processing*) sustentam processos de exploração expeditos para a realização de análises multidimensionais de dados, segundo as várias perspetivas de análises (as dimensões) integradas nesses *data warehouses*. As análises OLAP consistem,

basicamente, em explorar de forma interativa bases de dados multidimensionais, efetuando operações como *drilldown*, *rollup*, ou *slice e dice*, que são na realidade operações de detalhe, generalização, fatiamento (corte) e rotação (mudança de perspectiva).

Da mesma forma que surgiu a necessidade de introduzir a noção de preferência nos sistemas de base de dados, existe a mesma necessidade para os sistemas OLAP. A noção de preferência para sistemas base de dados tinha como objectivo satisfazer o consumidor na compra de produtos em sistemas de comércio electrónico. A sua motivação partiu do facto de se pretender responder nas aplicações, principalmente na Internet, de forma adequada ao utilizador, em situações em que este tenta encontrar um dado produto. Em sistemas OLAP a necessidade de expressar preferências do utilizador surge num contexto interno nas organizações, no qual os agentes de decisão procuram por informação capaz de suportar as suas decisões. Pretende-se aqui, sobretudo, com base no seu perfil de exploração focar o utilizador nos dados mais importantes e mais relevantes num dado contexto aplicacional.

A título de exemplo, e considerando um exemplo clássico da área de marketing, considere-se as funções dos utilizadores, a informação que consultam e o contexto e verifique-se que são diferentes entre si:

- O gestor de vendas está principalmente interessado nas receitas mensais, mas também costuma ver com mais detalhe os valores dessas receitas por semana quando estas excedem os 10 mil euros.
- O agente de vendas prefere observar os produtos que apresentam uma maior comissão de venda.
- O gerente do negócio está mais interessado em observar categorias de produtos no qual a quantidade de vendas foi baixa.

A partir daquilo que foi efectuado para os sistemas de base de dados, apareceu a necessidade de transpor esse conhecimento para os sistemas OLAP. A expressão de preferência para sistemas OLAP é um pouco mais exigente devido a vários factores, como por exemplo (Mouloudi, et al., 2006):

- A enorme quantidade de dados a serem manipulados.

- A complexidade dos dados provocada em grande parte pela existência de hierarquias entre os atributos.
- A complexidade das consultas devido à presença de operações de junção e de agregação.
- As diferentes formas de exibir a resposta a um utilizador.

As preferências têm maior expressividade nos sistemas OLAP, devido ao volume de informação envolvido e à forma como os utilizadores exploram a informação contida nas estruturas multidimensionais de dados. Durante uma sessão de exploração OLAP, um utilizador não sabe exactamente o que está a procurar, ele vai sucessivamente lançando questões, cujos resultados influenciam as questões seguintes, procurando razões que possam estar por detrás de determinados fenómenos. Estas sessões podem envolver um número muito elevado de questões e de resultados.

### **3.1.1 A Abordagem Quantitativa**

As primeiras investigações efectuadas no domínio OLAP seguiram a abordagem quantitativa, no qual as preferências são expressas segundo uma dada função de pontuação. Em (Stefanidis, et al., 2005) foi proposto um modelo com base na informação de um contexto que pode ser usada para caracterizar a situação de uma entidade (pessoa, lugar ou objecto) considerada relevante na interação entre o utilizador e aplicação. Este modelo de preferências assenta numa estrutura de dados multidimensional que, como sabemos, é designada usualmente por hipercubo ou, simplesmente por cubo (Figura 4).

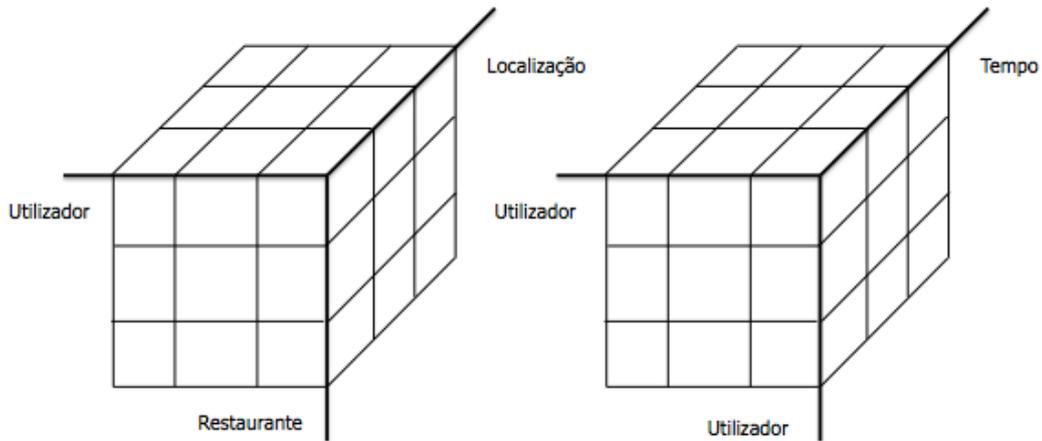


Figura 4 – Cubo de preferências apresentado em (Stefanidis, et al., 2005)

Abaixo é ilustrado um exemplo que pretende demonstrar as preferências segundo este modelo.

**Exemplo:** "A Maria toma sempre em consideração dois parâmetros de contexto na escolha de um restaurante: o tempo e a sua localização (Figura 4)."

De acordo com a estrutura, podemos definir uma preferência  $P1(187, 334, *, sol) = 0.8$ , ou seja, o restaurante "A Taberna" com o id = 187, para o utilizador "Maria" com o id = 334 tem um interesse de 0.8 quando está sol. O '\*' indica a ausência de valor (*null*). Neste caso a Maria não expressa nenhuma preferência quanto á localização. A Figura 5 revela a forma como é guardada a informação que expressa as preferências do utilizador.

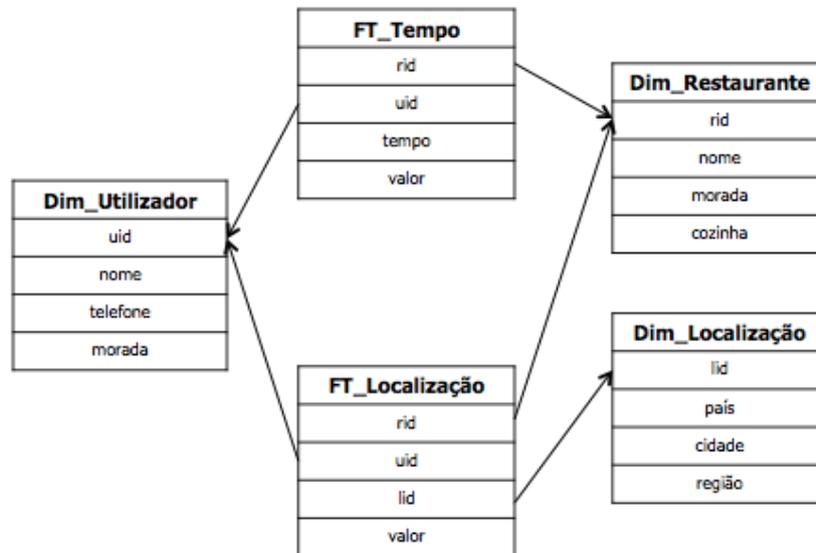


Figura 5 – Detalhe do modelo de preferências apresentado em (Stefanidis, et al., 2005)

O modelo proposto também possibilita a integração de preferências, ou seja, a possibilidade de fazer refletir duas ou mais preferências numa única (construção de preferências complexas).

### 3.1.2 A Abordagem Qualitativa

Apenas recentemente surgiu uma abordagem mais sólida para sistemas OLAP, cobrindo ambos os domínios, numérico e categórico, tal como aconteceu em (Kießling, 2002). Da mesma forma que nos sistemas de base de dados, Golfarelli e Rizzi (2009) estabeleceram as preferências sobre relações binárias (ordens parciais estritas). No domínio dos sistemas OLAP, as preferências dos utilizadores podem ser expressas sobre atributos e medidas, como também podem ser expressas ao nível da agregação de dados, indicando, por exemplo, que os dados mensais são preferíveis aos anuais ou aos diários. Golfarelli e Rizzi (2009) propuseram, também, uma álgebra para expressar as preferências dos utilizadores em sistemas OLAP, na qual as preferências complexas são definidas através da seguinte gramática:

```

< expr > ::= < baseConstr > | < expr > ⊗ < baseConstr >
< baseConstr > ::= POS | NEG | BETWEEN | LOWEST | HIGHEST |
CONTAIN | NEAR | COARSEST | FINEST

```

Esta gramática assenta num conjunto de operadores base (*baseConstr*) e no operador de composição Pareto ( $\otimes$ ). Os operadores base têm como objectivo expressar as preferências ao nível dos atributos, das medidas e das hierarquias.

### **Preferências Base ao Nível dos Atributos**

Estas preferências são aplicadas ao nível dos atributos, passando a explicar cada um dos casos possíveis:

- POS(*a*, *c*): Factos nos quais os valores *c* de determinado atributo *a* têm uma maior preferência sobre outros – ex. POS (Marca, 'BMW').
- NEG(*a*, *c*): Factos nos quais os valores *c* de determinado atributo *a* devem ser evitados – ex. POS(Marca, 'Fiat').

### **Preferências Base ao Nível das Medidas**

Estas preferências são aplicadas ao nível das medidas, passando a explicar cada um dos casos possíveis:

- BETWEEN(*m*, *v\_min* , *v\_max* ): Factos nos quais os valores entre *v\_min* e *v\_max* de determinada medida *m* têm preferência; para os restantes factos é estabelecido um ranking de acordo com a distância do intervalo de valores – ex. BETWEEN(Quantidade, 500, 1000).
- LOWEST(*m*), HIGHEST(*m*): Factos nos quais o valor para determinada medida *m* é mais baixo (ou alto) quanto possível – ex. LOWEST(Quantidade).

### **Preferências Base ao Nível das Hierarquias**

Estas preferências são aplicadas ao nível das hierarquias, passando a explicar cada um dos casos possíveis:

- CONTAIN(*h*, *a*): Factos agregados nos quais *a* está incluído têm preferência sobre os outros – ex. CONTAIN(Produto, Categoria).
- NEAR(*h*, *a\_fine*, *a\_coarse*): Factos agregados entre *a\_fine* e *a\_coarse* têm preferência sobre outros; para os restantes factos é estabelecido um ranking consoante a sua distância

- do intervalo de agregados – ex. NEAR(Tempo, Trimestre, Ano).
- COARSEST( $h$ ): Factos agregados sobre  $h$  têm preferência sobre os detalhados – ex. COARSEST(Produto).
  - FINEST( $h$ ): Factos detalhados sobre  $h$  têm preferência sobre os agregados – ex. FINEST(Produto)

### 3.2 Preferências com base em Ontologias

O objectivo da personalização na Web Semântica é tornar o acesso mais fácil aos recursos pretendidos. Poucos trabalhos de investigação têm sido apresentados no que diz respeito à expressão de preferências com base na Web Semântica. O trabalho proposto por Siberski (2006) obteve uma maior relevância nesta área, pois conseguiu mostrar como a noção de preferência de um utilizador é importante no contexto da Web Semântica. As linguagens semânticas permitem uma representação explícita da interação entre entidades, por exemplo, serviços, recursos e utilizadores, a um nível mais alto de abstracção, permitindo também o raciocínio automatizado sobre essa representação.

Na sua maioria as abordagens existentes (Siberski, 2006; Chamiel e Pagnucco 2008) tentam estender a linguagem de consulta SPARQL (Prud'hommeaux and Seaborne, 2006). Esta linguagem é utilizada para expressar consultas em diversas fontes de dados, nas quais estão guardados, nativamente, em *Resource Description Framework*<sup>8</sup> (RDF) ou são vistos em RDF através de um software intermédio. No entanto, existe um outro tipo de abordagem apresentada por Ayres e Furtado (2007) e por Toninelli, et al. (2008) que vai mais em conta com aquilo que é pretendido nesta dissertação e que consiste, em termos gerais, na representação do modelo de preferências com base numa ontologia.

Em (Siberski, 2006) é proposto uma extensão à linguagem SPARQL que suporte diretamente a expressão de preferências. Este trabalho teve como base o que foi efectuado em (Kießling, 2002) e em (Kießling e Kostler, 2002), devido às semelhanças existentes entre as linguagens SPARQL e SQL (Tabela 1).

---

<sup>8</sup> A linguagem RDF será apresentada, um pouco mais à frente, na secção 3.4.2.

Tabela 1 – Linguagem SPARQL e Linguagem SQL

Linguagem SPARQL	Linguagem SQL
<pre>SELECT &lt;projection-var-list&gt; FROM &lt;ontology-reference&gt; WHERE &lt;var-bindings&gt; FILTER &lt;hard-conditions&gt; ORDER BY &lt;var-list&gt; <b>PREFERRING &lt;soft-conditions&gt;</b></pre>	<pre>SELECT &lt;projection-list&gt; FROM &lt;table-reference&gt; WHERE &lt;hard-conditions&gt; <b>PREFERRING &lt;soft-conditions&gt;</b> ORDER BY &lt;attribute-list&gt;</pre>

Nesses trabalhos é ainda apresentada a aplicação prática da linguagem SPARQL recorrendo a um motor de consulta: o ARQ. Este motor de consulta realiza o seu trabalho em três fases: *Parsing*, *Query Planning* e *Query Execution*. Na segunda fase (criação do plano da query) é introduzida a noção de preferência na consulta. Chamiel e Pagnucco (2008) apresentaram também uma extensão à linguagem SPARQL, seguindo a mesma linha do que tinha sido feito por Siberski (2006) e recorrendo também as mesmas tecnologias na implementação.

O aumento da utilização de serviços em dispositivos móveis serviu de motivação ao trabalho de (Toninelli, et al., 2008) para conseguir melhorar a qualidade da informação que é visualizada pelo utilizador. Este trabalho pretendeu mostrar ao utilizador a informação, num determinado serviço, com base na sua localização e no dispositivo que ele está a utilizar. O perfil do utilizador também é tido em consideração, por exemplo, preferências, atividades, interações e ainda informação anexa a estas variáveis (e.g. data e hora das atividades e interações com serviços ou utilizadores). Neste trabalho é proposto uma camada intermédia de software designada por AIDAS (*Adaptable Intelligent Discovery of context-Aware Services*), que tem como base a Web Semântica para definição de uma ontologia de preferências, perfil de utilizador e serviços. Porém, esta abordagem sofre da falta de expressividade nas preferências de utilizador devido ao modelo básico que é apresentado. A Figura 6 apresenta o modelo ontológico de uma preferência base.

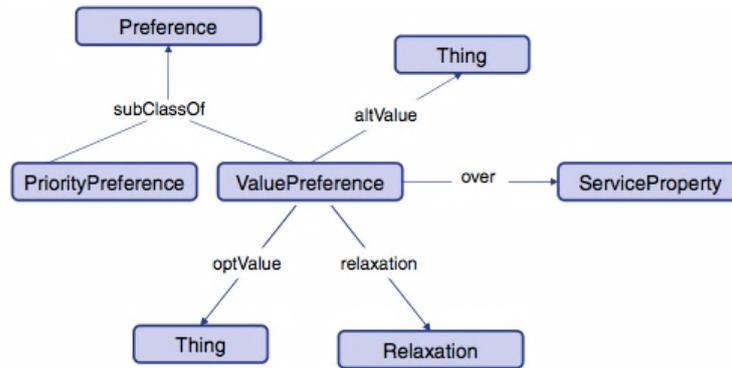


Figura 6 – Modelo ontológico de uma preferência base retirado de (Toninelli, et al., 2008)

Em (Tapucu, et al., 2009) foi um proposto um modelo formal de preferências de utilizador, genérico e partilhável, bem como uma extensão a uma linguagem de consulta. Este trabalho apresentou um estudo profundo à situação da expressão de preferências de utilizadores e da personalização de sistemas. O modelo proposto oferece um mecanismo de persistência e uma linguagem dedicada, sendo implementado utilizando bases de dados de base ontológica (OBDB). Uma OBDB é uma base de dados em que as ontologias e as instanciação dos dados estão juntas no mesmo repositório (Noy, 2004). O modelo de preferências proposto é formalmente definido utilizando a linguagem de modelação *EXPRESS* (ISO 10303-11).

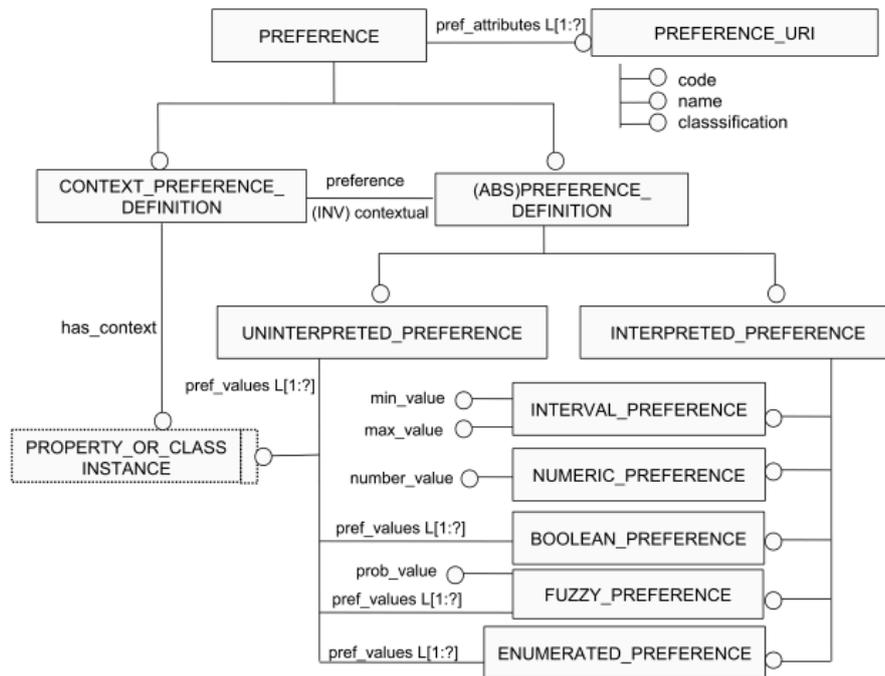


Figura 7 – Modelo de preferências retirado de (Tapucu, et al., 2009)

O modelo apresentado na Figura 7 revela a expressão de preferências de forma numérica, em intervalos de valores, booleano, *fuzzy* e enumerado. Nesse trabalho é ainda apresentada uma extensão à linguagem de consulta de ontologia OntoQL (Jean, et al., 2006), uma linguagem de consulta para bases de dados de base ontológica, que permite a expressão de consultas com a indicação das preferências do utilizador. O trabalho apresentado em (Tapucu, et al., 2009) tem como tema principal a expressão de preferências para sistemas de base de dados, tendo os autores apresentado alguns possíveis caminhos a seguir, de forma a dar continuidade ao modelo apresentado para lidar com sistemas de OLAP.

Tabela 2 – Estudo comparativo dos trabalhos realizados sobre preferências

<b>Autor</b>	<b>Domínio</b>	<b>Abordagem</b>	<b>Preferências</b>
Agrawal-Wimmers (2000)	Base de Dados	Quantitativa	Lógica
Börzsönyi, et al. (2001)	Base de Dados	Qualitativa	Lógica
Kiebling (2002-2003)	Base de Dados	Qualitativa	Lógica
Chomicki (2003)	Base de Dados	Qualitativa	Lógica

---

Stefanidis, et al. (2005)	Data Warehouse	Quantitativa	Lógica
Bellatreche, et al. (2005)	Data Warehouse	Qualitativa	Lógica
Mouloudi, et al. (2006)	Data Warehouse	Qualitativa	Lógica
Golfarelli and Rizzi (2009)	Data Warehouse	Qualitativa	Lógica
Siberski (2006)	Web Semântica	Ontológica	Semântica
Ayres e Furtado (2007)	Web Semântica	Ontológica	Semântica
Chamiel e Pagnucco (2008)	Web Semântica	Ontológica	Semântica
Toninelli, et al. (2008)	Web Semântica	Ontológica	Semântica
Tapucu, et. al. (2009)	Web Semântica	Ontológica	Semântica

### **3.3 Sumário**

Este capítulo revelou a evolução da expressão das preferências ao longo do tempo, tendo-se apresentado vários dos trabalhos que se foram realizando, destacando, quando possível, aqueles os que mais se evidenciaram. O processo de definição de preferências tem vindo a ser construído de forma incremental, evoluindo gradualmente ao longo dos anos. No entanto, ainda não existe uma definição concreta para preferência de utilização. Cada um tenta formular a sua própria definição e criar o seu próprio conceito, o que originou, obviamente diferentes tipos de abordagens. Num primeiro período fez-se a definição dos modelos e como estes se integravam nos sistemas, passando-se depois para um período de optimização.

A definição de preferências para os sistemas de base de dados está implementado, no entanto, a sua utilização ainda é pouco expressiva. Saliente-se o trabalho efectuado por Chomicki (2003) e por Kießling (2002), que representam um ponto de viragem na expressão de preferências em sistemas de base de dados. Quanto à expressão de preferências em sistemas OLAP, esta ainda está a passar por um processo de definição, não existindo por enquanto nenhum produto a ser comercializado que faça uso da noção de preferência. No entanto, já existem modelos e definições bastante sólidas para este domínio. Vejam-se, por exemplo os trabalhos de Golfarelli e Rizzi (2009).

Naquilo que diz respeito à Web Semântica ainda existe um longo caminho a percorrer, principalmente na utilização desta abordagem na personalização de sistemas OLAP. Neste contexto, Tapucu, et. al. (2009) realizaram o trabalho mais completo até ao momento, com o estudo que fizeram sobre as preferências de utilização e a personalização de serviços, mas

principalmente pela solução que apresentaram. Na Tabela 2 podemos ver as várias abordagens existentes, bem como os seus respetivos autores.

Dependendo da área de negócio e dos objectivos que se pretende atingir é utilizada uma das abordagens acima referidas. Tanto para os sistemas de base de dados como para os sistemas OLAP a falta de expressividade da abordagem quantitativa leva a que atualmente a sua utilização e as suas pesquisas tenham vindo a diminuir. Já o mesmo não acontece no que diz respeito à utilização da abordagem qualitativa e principalmente da Web Semântica.

## Capítulo 4

### Ontologias

As ontologias vêm preencher uma das lacunas que encontramos na Internet com o fornecimento de uma estrutura compartilhável e semântica para um determinado domínio. Desta forma desempenham um papel fundamental em diversas áreas de investigação, tais como o comércio electrónico ou o suporte à decisão. As ontologias são utilizadas para estudar a existência de todo o tipo de entidades (abstractas ou concretas) que constituem o mundo (Ceccaroni, 2001), fazendo hoje parte da Web Semântica.



Figura 8 – Web Semântica

As pessoas começaram a publicar cada vez mais conteúdos na Internet, desde simples interações com outras pessoas, através da criação de redes sociais (ex. Facebook e Twitter), passando pela partilha de conteúdo em blogs (ex. Blogger e Wordpress), a partilha de fotos e de vídeos (ex. Flickr e YouTube), até à contribuição de textos específicos em *Wikis* (ex. Wikipedia). A utilização de

*tags* e a leitura de conteúdos através de RSS *feeds* passou a ser uma prática comum. Também, surgiram novas tecnologias de desenvolvimento de aplicações para a Web, que conseguem oferecer mais e melhores conteúdos aos utilizadores. A esta "era" da Internet designou-se como Web 2.0. As páginas Web são escritas em HTML, que descreve e regula a estrutura e a apresentação da informação, ou seja, a sua sintaxe. Mas não a sua semântica.

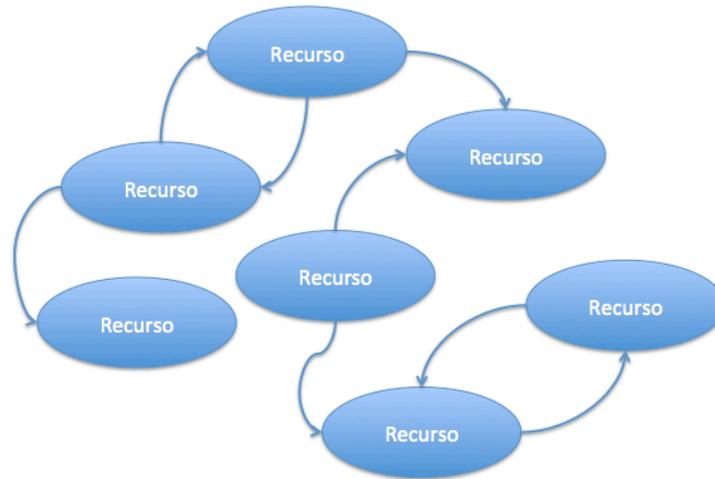


Figura 9 – Diagrama da Web Atual

As páginas são desenvolvidas com o único objectivo de serem vistas por pessoas. Não por máquinas. No entanto, se os computadores conseguissem perceber o significado da informação, poderiam aprender sobre o tipo de informação em que estamos interessados e ajudar de forma mais efetiva na sua procura. Esta "era" da Internet foi designada por Web 3.0. A Web atual é uma Web de documentos (Figura 9), enquanto que a Web Semântica é uma Web de "coisas" (Figura 10), muitas coisas.

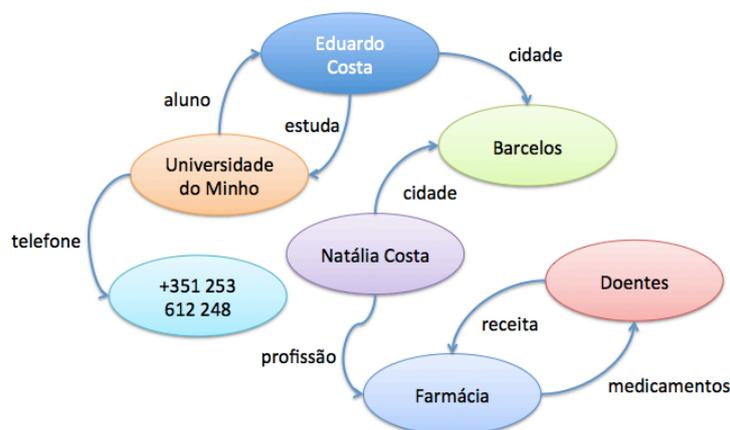


Figura 10 – Diagrama da Web Semântica

Depois da visão da Web Semântica, de realizar experiências, definir padrões para tecnologias, criar bibliotecas para o desenvolvimento de software e efetuar testes de conceito, a Web Semântica encontra-se hoje num estado já considerado de produção. Todavia, ainda há muito a fazer neste campo, principalmente no que diz respeito em conseguir obter mais disponibilidade das pessoas para aprender e construir soluções em Web Semântica. Relembre-se o que Tim Berners-Lee (1999) disse em tempos: "Eu tenho um sonho para a Internet em que cada computador seja capaz de analisar toda a informação na Internet."

## 4.1 Definição e Fundamentação

Do ponto de vista da Filosofia, ontologia é a teoria ou o estudo do ser, ou seja, das características básicas de toda a realidade. Apelidada no século 17, a ontologia é sinónimo de metafísica ou primeira filosofia, tal como tinha sido definido por Aristóteles no século 4 (Guarino, 1995). Mais recentemente, o termo foi adoptado por investigadores da área da Inteligência Artificial, que argumentaram a possibilidade de criar novas ontologias para a definição de modelos computacionais que permitissem certos tipos de raciocínio automatizados.

No início de 1990, foi desenvolvido um sério esforço para criar padrões de interoperabilidade, no qual a camada da ontologia foi designada como uma componente padrão dos sistemas de conhecimento. A especificação de termos e conceitos causou na altura muita polémica e deu

origem a vários debates, que fizeram emergir algumas conclusões relevantes, nomeadamente (Obitko, 2007):

- Uma ontologia especifica os conceitos, os relacionamentos e outras definições relevantes à modelação de um dado domínio.
- A especificação da ontologia assume as definições do vocabulário de forma representacional (classes, relações, etc.), obtendo-se significados para algumas limitações do vocabulário.

No âmbito das ciências da computação e da informação, uma ontologia define um conjunto de primitivas de representação que modelam um domínio de conhecimento ou um discurso. As primitivas de representação são tipicamente classes (ou conjuntos), atributos (ou propriedades) e relações (ou relações entre os membros da classe). Ontologias são tipicamente especificadas em linguagens que permitem uma captação longe das estruturas de dados e estratégias de implementação. A título de exemplo, no contexto dos sistemas de base de dados, a ontologia pode ser vista como um nível de abstração dos modelos de dados, análogo ao dos sistemas relacionais. Desta forma, as ontologias são vistas ao nível semântico, enquanto que as bases de dados são modelos de dados no nível lógico ou físico (Obitko, 2007).

As ontologias oferecem uma abstração para um problema, independente da sua forma particular. No exemplo acima referido, uma ontologia torna-se independente da tecnologia a usar, sendo esta uma das suas principais características. O nível de abstração também se pode tornar demasiado amplo, tornando-se assim mais difícil a sua percepção. No contexto das engenharias, as ontologias estão preocupadas em fazer a representação da informação mais relevante de um domínio, ao mais alto nível de abstração, não deixando de ser o mais claro possível sobre os significados dos termos. Atualmente, existem linguagens padrão e uma oferta muito diversificada de ferramentas comerciais (e não comerciais) para criar e trabalhar com ontologias.

## **4.2 Razões para desenvolver ontologias**

As ontologias têm um papel importante na partilha e reutilização do conhecimento. Em (Noy e McGuinness, 2003) são apresentadas algumas das principais razões para desenvolver uma ontologia, nomeadamente:

- *Partilhar conhecimento comum de estruturas de informação entre as pessoas ou agentes de software* é um dos objectivos mais comuns no desenvolvimento de ontologias. Considere-se, por exemplo, um conjunto de páginas Web com informações médicas ou que fornecem serviços de comércio electrónico. Se estas páginas Web partilharem e publicarem uma mesma ontologia dos termos que usam, então programas como os agentes de software podem extrair e agregar informação de vários pontos. Esta informação agregada pode ser utilizada para responder a consultas de utilizadores ou servir como uma base de dados para outras aplicações.
- *Permitir a reutilização de conhecimento* foi uma das grandes motivações para a investigação das ontologias. Por exemplo, no domínio do desenvolvimento de software existiu uma forte investigação para conseguir definir modelos capazes de conseguir expressar de forma abstracta todas os problemas que ocorriam no âmbito deste domínio. Tal investigação deu origem ao que hoje se designa por "Padrões de Desenho", que permitem aos que desenvolvem software seguirem normas de desenvolvimento e com isso conseguir reutilizar o que está a ser produzido em diferentes contextos.
- *Fazer suposições subjacentes a um domínio explícito da aplicação* torna possível alterar estes pressupostos facilmente se o nosso conhecimento sobre o domínio se alterar. Por exemplo, novamente no desenvolvimento de software, a escolha de uma linguagem de programação e a tecnologia a usar têm que ser independentes do modelo, uma vez que essa escolha é realizada para cada programador.
- *Separar o conhecimento do domínio do conhecimento operacional* é comum nas ontologias. Por exemplo, na construção e desenvolvimento de uma base de dados um dos primeiros passos a efetuar é o desenvolvimento do modelo relacional (ao nível do conhecimento do domínio), sendo efetuado o modelo lógico (ao nível do conhecimento operacional) apenas posteriormente, onde este já difere consoante a tecnologia de base de dados (ex. relacionais ou não relacionais).
- *Analisar o conhecimento do domínio* é possível uma vez que a especificação declarativa dos termos está disponível. A análise formal dos termos é extremamente valiosa quando é pretendido reutilizar ontologias de forma a conseguir de alguma forma estende-las (McGuinness, et al., 2000).

O desenvolvimento de uma ontologia é semelhante à definição de um conjunto de dados e da sua estrutura para permitir o uso de outros programas (uma API – *Application Programming Interface*).

### 4.3 Metodologias para o desenvolvimento de ontologias

Nos últimos anos tem havido um considerável progresso no desenvolvimento das bases conceituais para a construção de ontologias. Mesmo assim, ainda não existe uma forma ideal para desenvolver uma ontologia. Existem algumas metodologias nas quais se apresenta um conjunto de passos a seguir durante o processo de desenvolvimento, que é, usualmente um processo iterativo, lento e complexo. As metodologias que hoje mais se destacam são:

- *Enterprise Modeling (EM)*, que permite fazer a representação computacional da estrutura, atividades, processos, informação, recursos, pessoas, comportamentos, objetivos e restrições do negócio, do governo ou de outra empresa. Caracteriza-se por poder ser tanto descritiva como de definição. Basicamente, o objetivo desta metodologia é conseguir alcançar através do modelo desenvolvido, o desenho da empresa, dos seus processos de análise e das suas operações. Os passos que esta metodologia recomenda que se sigam são:
  - *A captura do domínio de conhecimento*, que requer que fique claro o porquê de se estar a desenvolver uma ontologia e qual a sua aplicação.
  - *A codificação*, que envolve representar explicitamente o conhecimento adquirido anteriormente numa linguagem formal.
  - *A integração com ontologias existentes*, no qual se valida se durante a captura ou durante a codificação dos processos, se deve usar ou não ontologias que já existem e de como o fazer no caso de as usar.
  - *A avaliação*, no qual se efetua uma análise técnica.
  - *A documentação*, passo em que se recomenda que todos os processos anteriores sejam documentados devidamente.
  
- *Toront Virtual Enterprise (TOVE)*, tem como base a EM. O grupo de desenvolvimento desta metodologia seguiu a abordagem da construção da EM baseado na motivação cenários, questões de competência formal e informal, especificação da terminologia e especificação

do axioma. A construção dos modelos da TOVE não é feita diretamente. Primeiro é realizada uma descrição informal das especificações a serem cumpridas pela ontologia, sendo essa descrição formalizada posteriormente. Esta metodologia apresenta os seguintes passos para o seu desenvolvimento:

- *Captura de cenários de motivação.* Os cenários motivadores são problemas ou exemplos que não são tratados adequadamente por ontologias existentes. Qualquer proposta de uma nova ontologia ou extensão a uma ontologia existente deve descrever um ou mais cenários de motivação, bem como o conjunto de soluções destinados aos problemas apresentados nos cenários.
  - *Questões de competência informal.* A ontologia deve ser capaz de representar estas questões utilizando a sua terminologia e ser capaz de lhes dar uma resposta utilizando os axiomas e definições.
- *Methontology*, que permite aos especialistas e aos criadores de ontologias que não estão familiarizados com o ambiente de implementação, construir ontologias desde o início. O aspecto mais distintivo desta metodologia é a sua manutenção. O ambiente de desenvolvimento para construir ontologias usa uma *framework* com o nome *Ontology Design Environment* (ODE). Esta metodologia apresenta os seguintes passos para o seu desenvolvimento:
    - *Especificação*, no qual se identifica o propósito e o âmbito da ontologia. O propósito deve responder a questões como “Porquê começar a desenvolver esta ontologia?”, enquanto que o âmbito deve responder a questões como “Qual é a sua finalidade e quais são os utilizadores finais?”.
    - *Conceptualização*, que representa num modelo conceptual a ontologia a ser construída, de forma a que este consiga responder às especificações.
    - *Formalização*, que é o processo que traduz a descrição conceptual num modelo formal, ou seja, a descrição do domínio estabelecido na etapa anterior é escrito de uma forma mais formal, não ficando contudo na sua forma final.
    - *Implementação*, passo no qual se faz a implementação da ontologia formalizada numa linguagem de representação do conhecimento.
    - *Manutenção*, no qual se faz a atualização e a correção da ontologia implementada.

- *101*, que é uma metodologia que tem como princípio base a existência de várias alternativas viáveis para desenvolver uma ontologia, que é um modelo da realidade do mundo no qual os seus conceitos devem refletir essa realidade. Esta metodologia utiliza a ontologia de domínio e âmbito (com base no conhecimento relacionado) para determinar, de uma forma pragmática, qual a melhor abordagem para o desenvolvimento de uma ontologia. A 101 desenvolvesse ao longo dos seguintes passos:
  - *Determinar o âmbito e o domínio da ontologia* através da resposta a algumas questões básicas como “Qual o domínio que a ontologia irá cobrir?”, “Que tipo de questões a ontologia deverá dar resposta?”.
  - *Considerar o uso de ontologias existentes*. Vale sempre a pena considerar o que outros já fizeram. Dessa forma é possível não começar completamente do zero e estender ou aperfeiçoar o trabalho realizado anteriormente.
  - *Enumerar os termos mais importantes*. Fazer uma lista dos termos de domínio mais importantes é muito útil, principalmente numa fase inicial já que permite obter uma primeira visão sobre o modelo.
  - *Determinar as classes e as hierarquias de classes*. Existem várias abordagens possíveis no desenvolvimento de uma hierarquia de classes: a abordagem de cima para baixo, a abordagem de baixo para cima e a combinação das duas. Nenhum dos três métodos é melhor do que o outro, tudo depende do ponto de vista pessoal do domínio.
  - *Definir propriedades*. As classes, por si só, não fornecem informação suficiente para responder às questões levantadas no passo 1. Cada classe tem um conjunto de propriedades que a caracterizam, ou seja, uma classe como Cliente contém atributos como nome, identificação, contacto, morada, etc.
  - *Definir restrições aos atributos*. Um atributo pode ser apresentado de várias formas: um texto, um número ou uma imagem.
  - *Criar instâncias*. Neste passo faz-se a criação de instâncias individuais de classes na hierarquia, ou seja, escolhe-se uma classe, cria-se uma instância individual dessa classe e preenche-se os seus atributos.

## 4.4 Linguagens de ontologias

Uma linguagem de ontologia é uma linguagem formal através do qual uma ontologia é construída. Existem várias linguagens para a implementação de ontologias, com diferentes características em termos de expressividade, facilidade de utilização e complexidade. As linguagens estão divididas, essencialmente, em três tipos:

- vocabulários definidos através da linguagem natural;
- linguagens baseadas em representação de objetos, como a UML;
- linguagens baseadas em predicados lógicos, como a lógica descritiva (Stevens, et al., 2000).

As linguagens mais clássicas foram desenvolvidas durante a década de 90, como por exemplo a KIF-based Ontolingua (Stanford University, 1997), a Loom (Loom, 2007) ou a Frame Logic (F-Logic) (Flora, 2008), que tem com base a lógica de primeira ordem ou a lógica descritiva. Mais recentemente surgiram outras linguagens, tais como a RDF (Beckett 2004), a RDFS (Brickley e Guha 2004), a XOL (Karp, et al. 1999), a SHOE (Helfin, et al., 1999), a OIL (OIL, 2000), a DAML+OIL (DAML, 2006a) ou OWL (McGuinness e Harmelen, 2004). A XML surge como standard na troca de informação na *Web*, que foi também adotada pelas linguagens para a implementação de ontologias.

### 4.4.1 Linguagens clássicas para a especificação de ontologias

Hoje temos à nossa disposição um conjunto muito diverso de linguagens que nos permitem fazer a especificação de uma dada ontologia. Desse conjunto salientamos a:

- *Ontolingua* é criada por volta dos anos 90 para suportar o desenho e a especificação de ontologias. O ambiente de desenvolvimento do Ontolingua oferece um conjunto de ferramentas para a criação de ontologias e uma biblioteca. É um serviço que está disponível na internet e tem uma comunidade de utilizadores considerável (Farquhar, et al., 1997).
- *Loom* foi desenvolvida ao mesmo tempo que a Ontolingua. É uma linguagem de representação de conhecimento baseada em lógica descritiva. Trata-se não só de uma linguagem, mas também de uma ferramenta para construção de aplicações “inteligentes” (Loom, 2007).

- *Operational Conceptual Modeling Language (OCML)* surge em 1993 e pode ser considerada como uma espécie de Ontolingua operacional, pois permite a especificação e a operacionalização de funções, relações, classes, instancias e regras. Inclui também mecanismos de definição de ontologias e métodos para a resolução de problemas (KMI, 2004).
- *F-Logic* foi desenvolvido em 1995 e combina frames<sup>9</sup> e lógica de primeira ordem. Esta linguagem é útil para criar ontologias básicas para integração de informação existente e conhecimento especializado.

#### 4.4.2 Linguagens de especificação de ontologias com base na Web

A internet promoveu a criação de ontologias com base na Web. O Consórcio World Wide Web (W3C) recomenda um conjunto de *standards* que fazem parte da estrutura da Web Semântica.

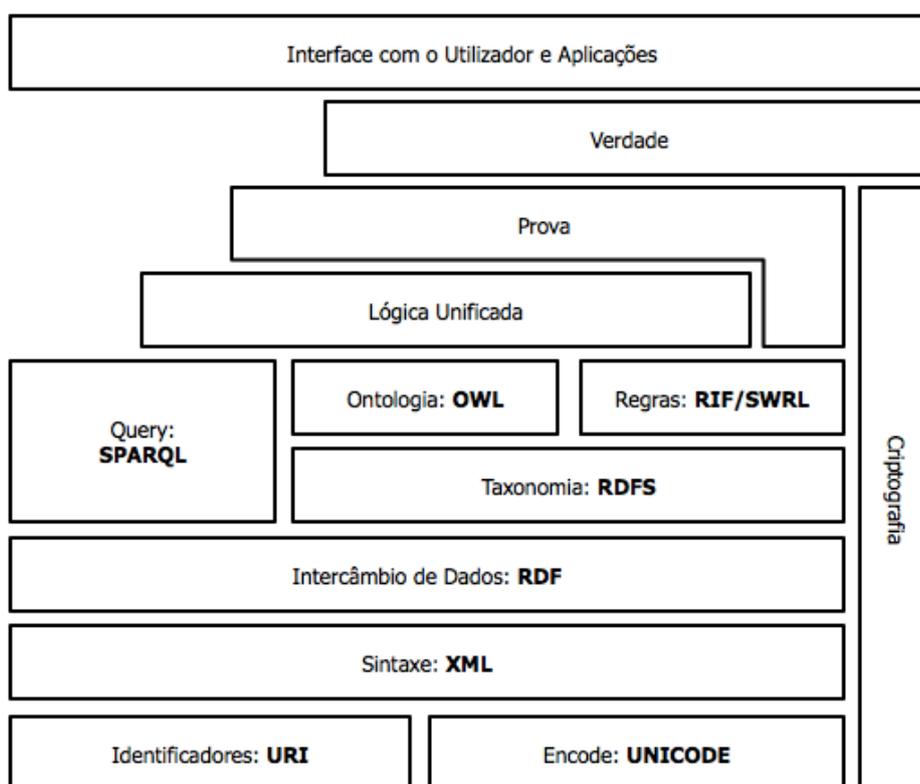


Figura 11 - Estrutura da Web Semântica – adaptado de (W3C, 2007)

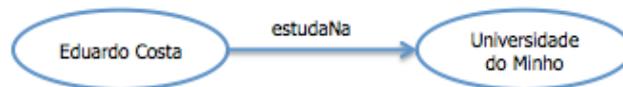
<sup>9</sup> Um *frame* consiste num objeto primitivo que representa uma entidade num determinado domínio.

AA Web Semântica é uma extensão da Web actual, que permite que os computadores e as pessoas trabalhem em cooperação. Surgiu em 2001 com o principal objectivo de desenvolver tecnologias e linguagens que tornem a informação legível para as máquinas (W3C, 2007).

A *OWL Web Ontology Language Overview* descreve a função e a relação de cada um dos componentes da Web Semântica (Stanford University, 1997):

- A *Extensible Markup Language (XML)* oferece uma sintaxe elementar para estruturar o conteúdo dentro de um documento. A XML Schema é uma linguagem para oferecer e restringir a estrutura e o conteúdo dos elementos contidos em documentos XML.
- A *Resource Description Framework (RDF)* é uma estrutura para representar a informação sobre os recursos em forma de grafo<sup>10</sup>, utilizando declarações simples (trio) para descrever as relações (ex. Sujeito-Predicado-Objecto). Veja-se os seguintes exemplos:

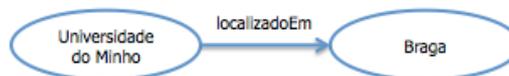
“Eduardo Costa estuda na Universidade do Minho”



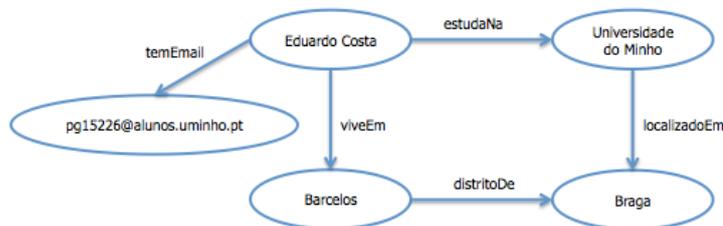
“Eduardo Costa tem o email pg15226@alunos.uminho.pt”



“O campos da Universidade do Minho está localizado em Braga”



<sup>10</sup> Grafo é o objecto básico de estudo da teoria de grafos. Tipicamente, um grafo é representado como um conjunto de pontos (vértices) ligados por retas (arestas).



- A *Simple Protocol and RDF Query Language (SPARQL)* é uma linguagem de consulta para fontes de dados da Web semântica. Esta linguagem permite:
  - Obter informação de dados estruturados e semi-estruturados.
  - Explorar dados através da descoberta de relacionamentos desconhecidos.
  - Consultar e pesquisar com uma visão integrada diferentes fontes de dados.
  - Reunir diferentes aplicações através da transformação de um vocabulário em outro.
- A OWL é uma extensão à RDF e tem como principal objectivo trazer o poder de expressão e raciocínio da lógica de descrição para a Web semântica. No entanto, nem tudo em RDF pode ser expresso em DL. A OWL adiciona mais algum vocabulário para descrever propriedades e classes: relações entre classes (ex: disjunção), cardinalidade (ex: "exatamente uma", igualdade e classes enumeradas).
- A *Semantic Web Rule Language (SWRL)* é uma linguagem com base em OWL. Permite aos utilizadores escrever regras que podem ser expressas em termos de conceitos OWL. Oferece uma melhor capacidade de raciocínio dedutivo quando comparada com o OWL.

## 4.5 Sumário

Já existe uma base sólida de conhecimento e um longo caminho percorrido na criação de modelos ontológicos. A metodologia de desenvolvimento que mais se destaca é a 101, enquanto ao nível de linguagens é a OWL que tem algumas vantagens em relação às outras e reúne, por agora, maior consenso à sua volta. Contudo, é necessário que mais pessoas entrem nesta temática de forma a dar o seu contributo com mais e melhores soluções. A falta de uma entidade reguladora para a Internet é algo que mais cedo ou mais tarde vai ter que ser resolvido, a forma descontrolada com esta está a crescer e a questão de como que a informação vai se organizar de num ambiente global são alguns dos aspetos que vão precipitar tal necessidade.

## **Capítulo 5**

### **Uma Ontologia para Preferências**

A partilha de informação é cada vez mais uma necessidade real, sempre que nos deparamos com mais informação em cada um dos problemas que enfrentamos. O crescente progresso das tecnologias de redes e armazenamento de informação permitem hoje a digitalização e a difusão de grandes quantidades de documentos. Muito trabalho tem sido feito de forma a conseguir a comunicação entre sistemas heterogêneos, nos quais se destacam, a título de exemplo, os relacionados com a área da saúde.

A título de exemplo, hoje em dia existe a possibilidade de acesso às nossas contas bancárias, fotografias e tarefas especificadas em calendários disponíveis da Internet, e muitas outras coisas mais. No entanto, e como curiosidade, não é possível visualizar muitas vezes o dia em que as fotografias foram tiradas ou o que se fez nesse dia e o valor monetário que gastámos com elas. Essa informação também, muitas vezes, está disponível na Internet, só que é controlada de forma independente por diferentes aplicações e de diferentes maneiras. Devido aos modelos de ontologias padrão existentes (e.g. OWL, RDF, entre outras), várias ontologias têm sido criadas, aumentando conseqüentemente a quantidade de dados descrita por ontologias.

Na maioria dos sistemas de informação existentes nas organizações, as preferências do utilizador não são modeladas explicitamente. Estas são normalmente inseridas de forma manual e disseminadas através de aplicações que exploram esses sistemas. Como consequência, não podem

ser partilhadas entre os vários sistemas que um utilizador explora. Usualmente, as preferências do utilizador são definidas e atualizadas por cada aplicação.

Neste capítulo é apresentado o modelo para uma ontologia que permita expressar as preferências de utilizadores em sistemas OLAP, bem como se expõe como esse modelo é constituído e se explica o seu processo de desenvolvimento. A ontologia desenvolvida caracteriza-se por três componentes principais que consiste, nomeadamente: no perfil de utilizador, na expressão das preferências OLAP e no esquema de base de dados multidimensional. O desenvolvimento da ontologia seguiu a abordagem *top-down*, apresentada na metodologia 101, que compreende numa primeira fase a seleção dos conceitos gerais mais importantes e depois na especialização de cada um deles.

## 5.1 Definição do Domínio e do Âmbito

A definição do domínio da ontologia e a criação de um conjunto de interrogações no qual a ontologia deve responder é uma fase importante do seu desenvolvimento. Pois, durante o seu desenvolvimento é possível interrogar a ontologia, de forma a conseguir verificar se está de acordo com o que tinha sido especificado anteriormente. A resposta a algumas das questões básicas apresentadas pela metodologia 101 ajuda a perceber qual o domínio que esta ontologia vai cobrir e perceber como é que será efetuada a sua utilização. Vejamos, então, um exemplo concreto desse processo de pergunta-resposta:

- P:** Qual o domínio que esta ontologia vai cobrir?
- R:** Preferências de utilizadores em sistemas OLAP.
- P:** De que forma é que vai ser utilizada a ontologia?
- R:** A ontologia será utilizada como fonte e partilha de informação sobre as preferências de utilizadores entre sistemas, de forma a conseguir uma melhor interatividade entre o utilizador e as estruturas multidimensionais – hipercubos - envolvidas.
- P:** Que tipo de questões a ontologia deve dar resposta?
- R:** A ontologia deve responder a questões como:
- Quais as características de determinado utilizador?
  - Quais as preferências de cada utilizador?
  - Quais as consultas efectuadas por determinado utilizador?

- P:** Quem vai utilizar e manter a ontologia?
- R:** A ontologia deverá ser entendida como uma base de conhecimento onde a informação é proveniente de diversas aplicações de diferentes domínios. A sua utilização como fonte de conhecimento deverá ser efetuada por todos aqueles que pretendam ver melhorada a navegação do utilizador numa estrutura de dados multidimensional, tendo em atenção as suas preferências.

As respostas a estas questões pode sofrer alterações durante o processo de desenvolvimento da ontologia, uma vez que têm que acompanhar o próprio domínio da ontologia. No desenvolvimento da ontologia podem surgir também alguns problemas que obriguem ao repensar do seu domínio ou até mesmo estendê-lo para que possa cobrir novos cenários.

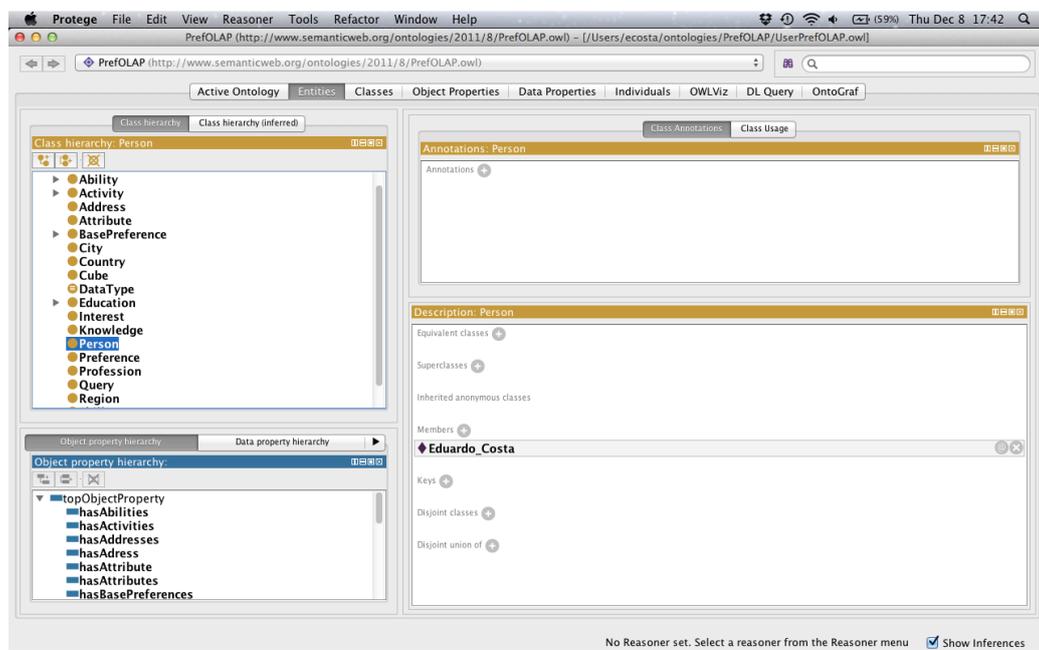


Figura 12 – O ambiente do Protégé para o desenvolvimento de uma ontologia

## 5.2 O Processo de Desenvolvimento

Tal como referido anteriormente, o processo de desenvolvimento desta ontologia seguiu a metodologia 101. Durante este processo existiu uma atenção especial para que fosse possível a sua reutilização, não só como um todo como também de cada um dos componentes, visto que a

ontologia em questão é caracterizada por 3 componentes distintas: perfil de utilizador, modelo multidimensional e preferências de utilização. O *software* usado no desenvolvimento da ontologia foi o Protégé [9], que oferece um conjunto de funcionalidades bastante versáteis que ajudam no desenvolvimento, verificação e validação da ontologia a desenvolver. A Figura 12 apresenta o ambiente de desenvolvimento de uma ontologia disponibilizado pelo Protégé.

### **5.2.1 Determinar o âmbito e o domínio da ontologia**

A determinação do âmbito e do domínio de uma ontologia foi estabelecido com a resposta às questões apresentadas pela metodologia 101, definida na parte relativa à "Definição do Domínio e do Âmbito".

### **5.2.2 Considerar o uso de ontologias existentes**

Embora seja aceite que se construam ontologias a partir do zero, o desenvolvimento de novas ontologias não consegue retirar todo o potencial de fontes de conhecimento eventualmente já existentes em domínios considerados relevantes para o problema em causa. A reutilização de ontologias pode ser definida como um processo em que o conhecimento (ontológico) disponível é utilizado como entrada para gerar novas ontologias. É um processo recomendado nas metodologias atuais, sendo considerado um factor chave para desenvolver ontologias de forma rentável e de qualidade. A reutilização de ontologias é um processo que deve reduzir os custos de engenharia, uma vez que evita a reconstrução de ontologias existentes. Como em qualquer outra disciplina de engenharia, a reutilização de alguns componentes implica o custo de pesquisa para os encontrar, de os conseguir perceber e, por fim, de os adaptar a um novo contexto. Por vezes a reutilização tem custos e implica esforços que podem ser mais elevados do que aqueles que estariam envolvidos num processo de construção a partir do zero (Bontas, et al., 2005). Existem muitas fontes de ontologias disponíveis na Internet. Por exemplo:

- Sharable Ontologies Library (Gruber, 2000).
- DAML Ontology Library (DAML, 2006b).
- Gene Ontology (GO, 1999).

Durante a fase de investigação sobre o desenvolvimento de ontologias, despendemos alguns esforços na procura de ontologias dentro da temática deste trabalho. Com esse processo

conseguimos encontrar algumas ontologias relativas ao perfil de um utilizador (Golemati, et al., 2007) e também encontrar alguns trabalhos de modelos de ontologias que expressam as preferências dos utilizadores (Ngoc, et al., 2005). Tal como já referido, no que diz respeito a ontologias que representem um esquema multidimensional, não foram encontrados trabalhos com relevância para o pretendíamos realizar no âmbito desta dissertação.

Neste trabalho foi efectuada uma reutilização quase na sua totalidade de (Golemati, et. al., 2007) em termos do perfil de um utilizador, tem sido realizadas somente algumas pequenas alterações de forma a conseguir a sua integração no caso particular da ontologia que desenvolvemos. Quanto às restantes duas componentes da ontologia desenvolvida, apenas foram retiradas algumas ideias de outros trabalhos, pois o seu conteúdo não se ajustava ao contexto do trabalho que tínhamos em curso.

### **5.2.3 Enumerar os termos mais importantes**

A enumeração dos termos mais importantes numa lista é um processo bastante útil, pois permite ao utilizador obter um primeiro esboço do seu modelo. Esta lista permite ao utilizador aprofundar cada um dos termos individualmente, tornando o processo de definição da ontologia mais rápido e eficaz. Depois, o estabelecimento das conexões entre as classes torna-se mais simples, no caso de se utilizar apenas as classes mais importantes. Mas, mesmo assim, é uma ajuda importante. A título de exemplo, numa ontologia para vinhos:

- Vinho, Uva, Adega, Localização.
- Cor do Vinho, Corpo do Vinho, Sabor do Vinho, Teor de Açúcar.
- Vinho Branco, Vinho Tinto, Vinho Rosé.

Na ontologia de preferências, termos como 'Pessoa', 'Preferência' e 'Cubo OLAP' são aqueles que mais se evidenciam. A especificação dos restantes termos passou por pegar em cada um destes e fazer o seu detalhe. Isto é 'Actividades', 'Interesses' e 'Skills' são alguns termos importantes na definição do perfil de utilizador.

#### 5.2.4 Determinar as classes e as hierarquias de classes

A abordagem utilizada no desenvolvimento da hierarquia de classes foi claramente uma abordagem de cima para baixo (*top-down*), que de acordo com o domínio em questão pareceu ser a mais indicada. Uma classe resulta de um conceito para um determinado domínio é um conjunto de elementos com propriedades similares, por exemplo:

- Uma classe de *Vinhos*.
- Uma classe de *Adegas*.
- Uma classe de *Vinhos Tintos*.

Para o nosso caso, teremos então classes como:

- Uma classe de *Pessoas*.
- Uma classe de *Interesses*.
- Uma classe de *Actividades*.
- Uma classe de *Preferências*.

Por vezes, determinar as classes e as suas hierarquias torna-se uma tarefa complicada, pois é comum existirem dúvidas entre aquilo que é uma "classe ou instância" e aquilo que é uma "classe ou propriedade". Para isto, a solução passa muito pela sensibilidade de quem está a desenvolver a ontologia e pela verificação de qual é a melhor solução no âmbito da aplicação. É comum, também, por vezes, existir demasiado detalhe na ontologia. Novamente, tudo depende da sensibilidade de quem está a desenvolver em saber parar e ver até que ponto a aplicação precisa ou não desta ou daquela definição. A definição de hierarquias é muito importante não só porque permite uma visualização mais fácil, mas principalmente pela disponibilização da herança de propriedades presente no desenvolvimento de ontologias.

#### 5.2.5 Definir propriedades

Tal como referido anteriormente, as propriedades são os atributos/características de uma classe. No exemplo da classe 'vinhos', a 'idade', a 'cor' ou o 'volume de álcool' são características de um vinho. Da mesma forma que no processo anterior, é aconselhável fazer apenas a descrição de

características no âmbito da aplicação em causa, já que o contrário pode tornar-se num processo demasiado demorado devido à grande quantidade de características envolvidas.

No nosso caso, podemos dizer que, a definição de propriedades foi muito básica, pois apenas foi efectuado um processo de detalhe de alto nível, não tendo sido realizado nenhum desses processos para as classe. Propriedades como 'nome', 'idade' e 'altura' para a classe 'pessoa' não foram definidas, pois numa fase inicial (*proof of concept*) tal processo não vem acrescentar nada ao modelo, simplesmente porque o domínio não o exige.

### **5.2.6 Definir restrições aos atributos**

Faz parte das boas práticas de desenvolvimento de ontologias a restrição dos valores dos atributos, como também a definição dos seus valores por omissão. Desta forma o processo de instanciação fica mais facilitado, sendo assim também mais fácil perceber quais os tipos de valores que cada atributo pode assumir. Quanto à restrição de propriedades, esta é efectuada sobretudo relativamente ao tipo dos atributos (texto, numérico, booleano, instância de classe, etc.), aos valores permitidos, à sua cardinalidade, e à delimitação do intervalo de valores, se o atributo for de preenchimento obrigatório. Tal como a definição de propriedades, a restrição de atributos não foi um processo que teve grande ênfase, uma vez que nesta fase de prova de conceito tal não se justifica.

### **5.2.7 Criar instâncias**

A criação de instâncias é o último passo no desenvolvimento de uma ontologia. Todavia, não é o menos importante. A instanciação de classes permite validar o modelo da ontologia, de forma a verificar se este cumpre ou não com os requisitos estabelecidos. A instanciação da ontologia será apresentada mais tarde no capítulo 6, no qual será demonstrado com exemplos práticos concretos o modelo de ontologia desenvolvido e de seguida discutido.

## **5.3 O Modelo Desenvolvido**

Nesta secção apresentamos o modelo de ontologia desenvolvido, detalhando cada um dos componentes que o constituem. O modelo não pretende descrever de forma exhaustiva cada

propriedade das classes estabelecidas, mas especificar apenas o necessário para se criar uma base sólida que permita o constante desenvolvimento do modelo no âmbito das preferências dos utilizadores na exploração de sistemas OLAP.

### 5.3.1 Caracterização de um Perfil de Utilizador

A utilização do perfil do utilizador para a personalização de serviços é comum e tem como objectivo conseguir obter informação útil para expressar os seus gostos. A utilização de ontologias para modelar o perfil de utilizador já foi proposta em várias aplicações, como em processos de pesquisa na Internet (Trajkova et al., 2004) (Lawrence, 2000) e na gestão de informação pessoal (Katifori et al., 2005). No entanto, os modelos de ontologia propostos nesses trabalhos foram desenvolvidos para um domínio de aplicação específico. Todavia, em (Golemati et al., 2007) é proposto um modelo de ontologia genérico, com a importante característica de ser extensível, o que permite que o modelo possa ser adaptado às necessidades de cada aplicação, mantendo ao mesmo tempo uma estrutura geral comum capaz de satisfazer as exigências da portabilidade e da comunicação entre diferentes aplicações.

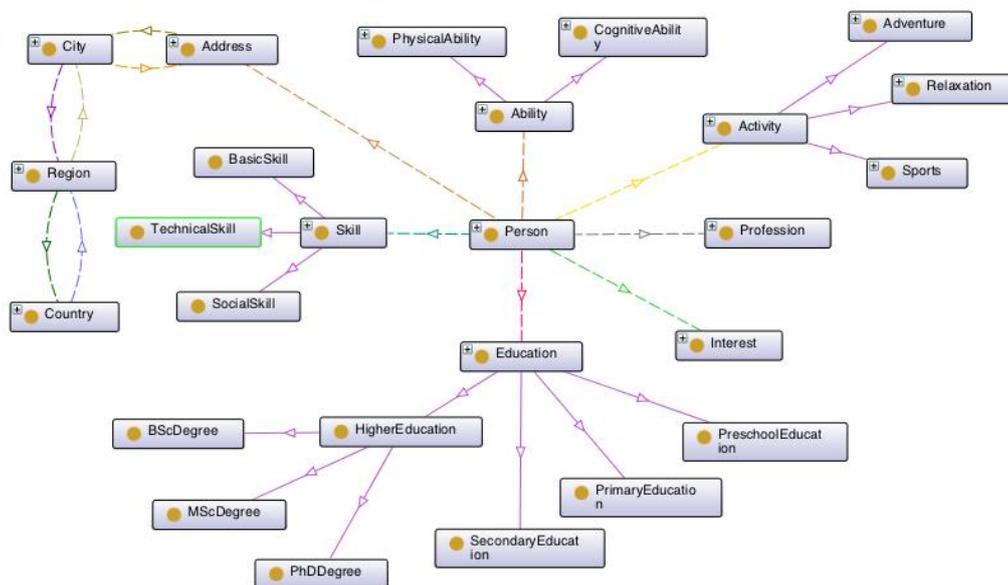


Figura 13 – Modelo de ontologia para o perfil de utilizador

Nesta secção apresenta-se uma breve descrição da ontologia que representa o perfil do utilizador. O trabalho de (Golemati et al., 2007) serviu como base de trabalho para o modelo desenvolvido

(Figura 13). A ontologia foi concebida de forma a que fosse possível facilmente adicionar novas classes que ajudem a caracterizar melhor um utilizador. Isso será importante para outros domínios de aplicação. Além disso, esta ontologia apresenta informação que é estática e permanente. Características mais dinâmicas como a posição atual do utilizador ou como a sua disposição está num determinado momento não estão a ser tratadas. No entanto, a ontologia permite a criação de novas instâncias de classes que representem características que possam mudar ao longo do tempo.

A Tabela 3 apresenta uma breve descrição para cada uma classes presentes no modelo de ontologia para o perfil de utilizador, de forma ser melhor compreendido.

Tabela 3 – Classes do modelo do perfil de utilizador

Classe	Descrição
Person	Informação básica do utilizador, como o nome, data de nascimento e e-mail.
Ability	Capacidades do utilizador, tanto ao nível físico como cognitivo.
Activity	Atividades e <i>hobbies</i> do utilizador. Estão divididas em aventura, relaxamento e desporto.
Address	Informação relativa à localização onde o utilizador habita.
Education	Educação obtida pelo utilizador, como diploma obtidos e formações efetuadas. Estão divididas em educação pré-primária, primária, secundária e universitária.
Interest	Interesses do utilizador, como "interesse em jogos, tecnologia e música". Também abrange interesses relacionados com o trabalho.
Skill	Habilidades do utilizador, como programar ou <i>design</i> . Estão divididas habilidades básicas, técnicas e sociais.

A classe "Person" é o centro de toda a ontologia, pois através dela são estabelecidas relações com as restantes classes que definem todas as características do utilizador. A Figura 14 apresenta um exemplo de um perfil de utilizador. Pela ilustração apresentada podemos retirar que o 'Eduardo Costa' vive na 'Rua Vasco da Gama nº 40', tem como educação uma licenciatura e um mestrado. Pratica como atividades 'BTT', 'geocaching', 'ténis' e 'futebol'. Tem também boas capacidades de 'comunicação', tanto ao nível da 'expressão' como da 'compreensão' e tem 'originalidade' naquilo que faz, lidando com facilidade com números e possui alguma força fisicamente. É uma pessoa interessada por 'música' e 'tecnologia'. Como especialidades 'Eduardo Costa' apresenta a 'reparação', a 'programação' e o '*design*', estando sempre atento ao ambiente que está envolvido.

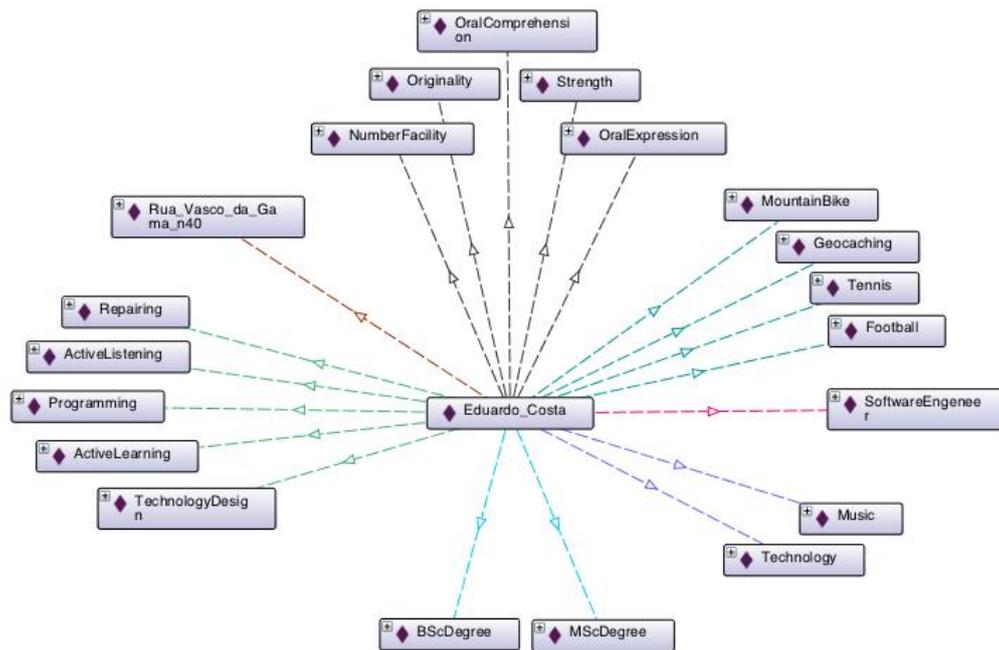


Figura 14 – Exemplo de um perfil de utilizador

### 5.3.2 Um Esquema Multidimensional

Como sabemos, a modelação de um esquema multidimensional é bastante complexa. Consequentemente, não existem muitos trabalhos que abordem, de forma explícita, este componente. De certa forma, e tanto quanto pudemos perceber, este trabalho torna-se pioneiro na definição de um modelo de ontologia capaz de suportar a definição de um esquema multidimensional. Esta secção apresenta o modelo desenvolvido, como também apresenta um pequeno exemplo prático ilustrativo da sua utilização.

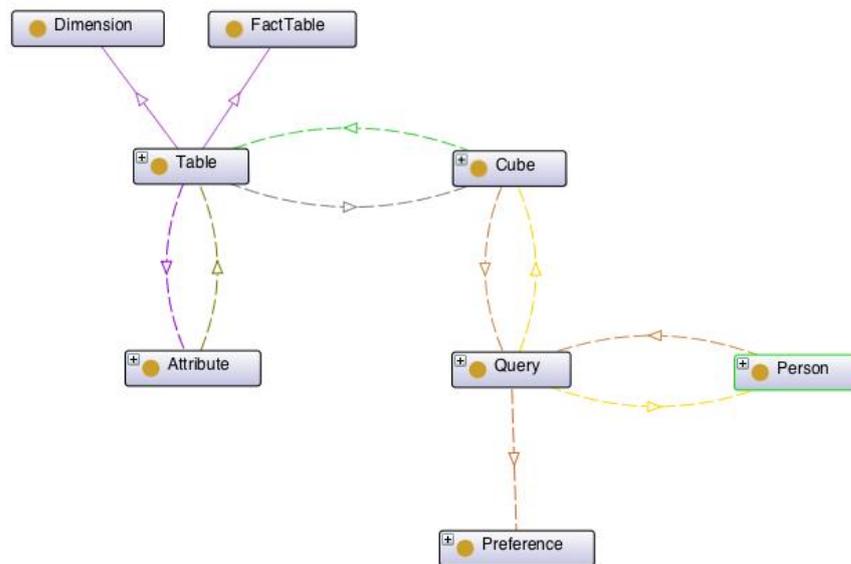


Figura 15 – O modelo de ontologia para o esquema multidimensional

Este modelo (Figura 15) tem não só como objectivo suportar o esquema multidimensional de uma base de dados, mas também fazer a ligação com as consultas efetuados pelos utilizadores e as respectivas preferências presentes nessas *queries*. A Tabela 4 apresenta uma breve descrição para cada uma classes presentes no modelo de ontologia para o esquema multidimensional, de forma ser melhor compreendido.

Tabela 4 – Classes do modelo do esquema multidimensional

Classe	Descrição
<b>Cube</b>	Informação básica sobre o cubo, como nome e descrição.
<b>Table</b>	Conjunto de dados dispostos em número finito de colunas e número ilimitado de linhas (ou tuplos).
<b>Dimension</b>	Em um cubo, uma dimensão é um elemento de dados que categoriza cada item em um conjunto de dados em regiões não sobrepostas.
<b>Fact Table</b>	Uma tabela de factos é composta das medições, métricas ou factos de um processo de negócio. Muitas vezes, está localizada no centro de um esquema em estrela ou de um esquema em floco de neve, onde é rodeada por tabelas de dimensão.
<b>Attribute</b>	Os atributos são os campos da tabela (dimensão / tabela de factos), e caracterizam os tipos

	de dados que deverão constar na tabela (numéricos, alfanuméricos, datas, coordenadas, etc.).
<b>Query</b>	Caracteriza uma consulta efetuada ao sistema de base de dados.
<b>Preference</b>	Consiste na definição de um filtro para uma determinada consulta efetuada ao cubo, de forma a conseguir satisfazer melhor o utilizador. É constituída por um conjunto de preferências base.
<b>Person</b>	Informação básica do utilizador, como o seu nome, data de nascimento e e-mail.

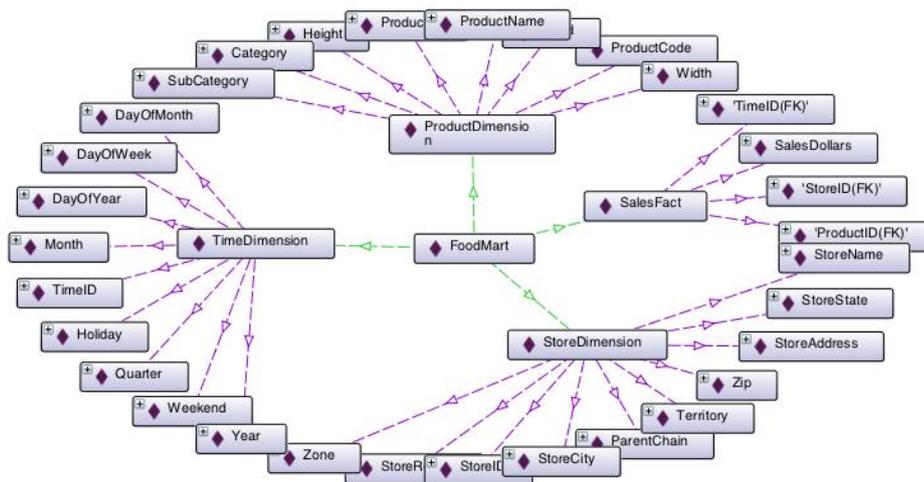


Figura 16 – Exemplo da especificação de um esquema multidimensional

A Figura 16 apresenta um exemplo da especificação ontológica de um esquema multidimensional. O esquema (Anexo A) é constituído por três dimensões: ProductDimension, StoreDimension e TimeDimension; e por uma tabela de factos: SalesFact. As dimensões e a tabela de factos integram, respetivamente, os seguintes conjuntos de atributos:

- ProductDimension – {ProductID, ProductName, ProductCode, Category, SubCategory, Brand, Width, Height}
- StoreDimension – {StoreID, StoreName, ParentChain, StoreRegion, StoreState, StoreAddress, Zone, Terretory, Zip}
- TimeDimension – {TimeID, Year, Quarter, Month, DayOfWeek, DayOfMonth, DayOfYear, Weekend, Holiday}
- SalesFact – {ProductID, StoreID, TimeID, SalesDollars}

### 5.3.3 Reflexo das Preferências OLAP

Tanto quando é do nosso conhecimento, a modelação das preferências de um utilizador não tem sido um alvo de estudo preferencial, principalmente no que diz respeito a sistemas OLAP. O modelo desenvolvido nesta dissertação pretende representar as preferências dos utilizadores desses sistemas, seguindo a abordagem apresentada por Golfarelli e Rizzi (2009).

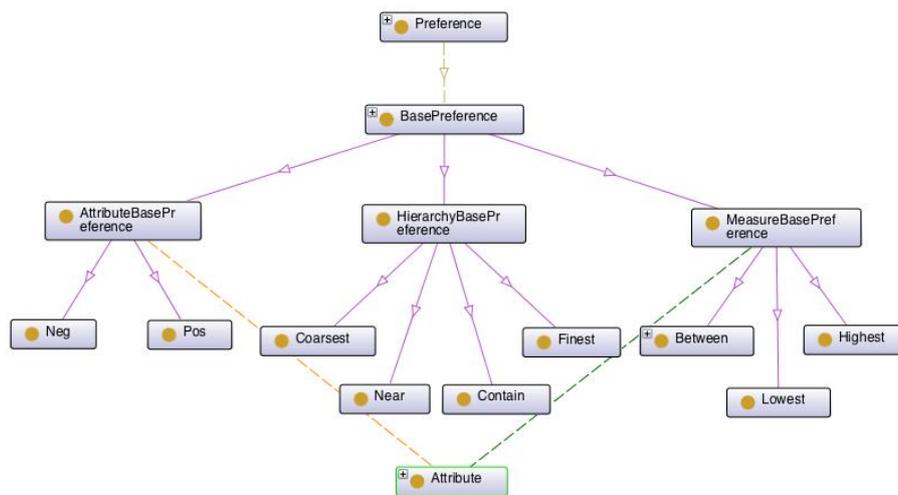


Figura 17 – O modelo de ontologia para preferências OLAP

Este modelo é uma primeira abordagem ao problema, servirá como ponto de partida para uma definição mais concreta. A priorização de preferências e a expressão de preferências ao nível das hierarquias é algo complexo e que este modelo não cobre na sua totalidade. A Tabela 5 apresenta uma breve descrição para cada uma das classes presentes no modelo de ontologia para as preferências OLAP, de forma a ser melhor compreendido.

Tabela 5 – Classes do modelo de preferências

Classe	Descrição
Preference	Consiste na definição de um filtro para uma determinada consulta efetuada ao cubo, de forma a conseguir satisfazer melhor o utilizador. É constituída por um conjunto de preferências base.
Base Preference	As preferências base podem ser de três tipos (preferências sobre os atributos, hierarquias ou medidas).

<b>Attribute Base Preference</b>	As preferências sobre os atributos podem ser definidas de duas formas (POS, NEG), onde o utilizador estabelece se tem preferência ou não por um determinado conjunto de valores sobre um atributo.
<b>Hierarchy Base Preference</b>	As preferências sobre as hierarquias podem ser definidas de quatro formas (Coarsest, Finest, Near e Contain), onde o utilizador estabelece a sua preferência sobre uma hierarquia em específico de forma a obter maior ou menor detalhe sobre uma determinada hierarquia na sua consulta.
<b>Measure Base Preference</b>	As preferências sobre as medidas são definidas de três formas (Highest, Lowest e Between), onde o utilizador estabelece a sua preferência num intervalo de valores, por valores elevados ou baixos sobre uma medida em específico.

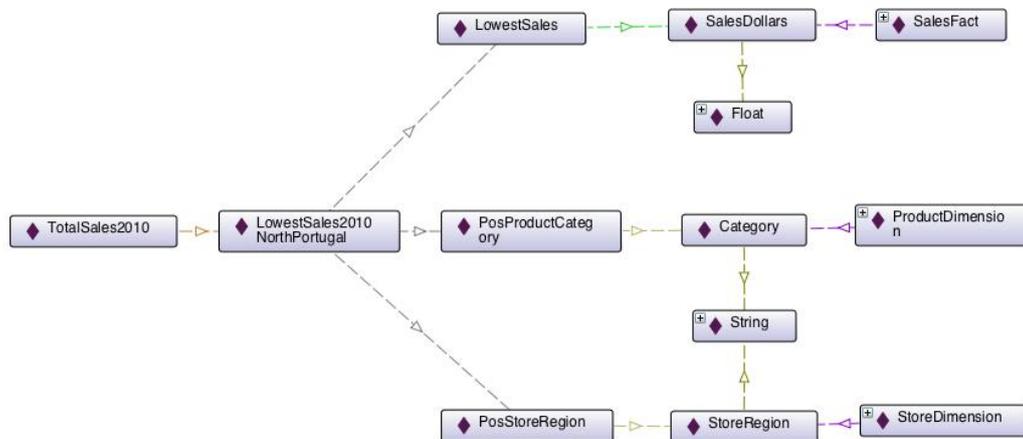


Figura 18 – Exemplo de uma preferência

A Figura 18 apresenta um exemplo de uma preferência de um utilizador ao executar uma consulta para saber qual o total das vendas realizadas em 2010.

### Query

```
SELECT SUM(SalesFact.SalesDollars) AS SumOfSalesDollars
FROM TimeDimension INNER JOIN (ProductDimension INNER JOIN
    SalesFact ON ProductDimension.ProductID =
    SalesFact.ProductID)
    ON TimeDimension.TimeID = SalesFact.TimeID
WHERE TimeDimension.Year=2010
```

**Preferência**

POS (StoreRegion, {Braga, Viana do Castelo, Porto}) ⊗ POS (Category, {carne, peixe}) ⊗ LOWEST (SalesDollars)

## **5.4 Sumário**

O desenvolvimento do modelo para a ontologia foi realizado sem dificuldades de maior, uma vez que existiu um estudo prévio relativamente ao desenvolvimento de novas ontologias, como também sobre qual a melhor metodologia a seguir e quais as ferramentas a utilizar. A consulta e o estudo de outras ontologias também contribuiu para tornar este processo de desenvolvimento mais simples. À medida que os componentes foram desenvolvidos (perfil de utilizador, esquema multidimensional e preferência OLAP) foram criadas algumas instâncias de forma a realizar a sua validação.



## **Capítulo 6**

### **O Caso de Estudo**

#### **6.1 Apresentação e Caracterização do Problema**

Para ilustrar o trabalho desenvolvido na concepção da ontologia, vamos considerar a existência de alguns personagens, nomeadamente: o 'João Carvalho', a 'Raquel Cruz', o 'Pedro Sousa' e o 'Miguel Lopes', todos eles agentes de decisão de uma empresa da área do retalho. A empresa opera unicamente em Portugal, no entanto está distribuída por todo o país com 82 pontos de venda. Atualmente encontra-se em forte crescimento, o que provocou um aumento significativo do volume de dados, criando, conseqüentemente, a necessidade de desenvolver novas técnicas e novas ferramentas que lhe permitam melhorar o trabalho dos seus agentes de decisão. Quanto à competências dos personagens referidos, podemos dizer que o 'João Carvalho' é o responsável pelos 'Refrigerantes e Vinhos' na 'Zona Sul', a 'Raquel Cruz' é responsável pela área dos 'Congelados', e o 'Pedro Sousa', juntamente com o 'Miguel Lopes' pelos 'Legumes e Frutos'.

Nas próximas secções iremos demonstrar de que forma a utilização de modelos ontológicos, nomeadamente um modelo de preferências, permite a partilha de informação entre os vários intervenientes e ajudar na criação de preferências de forma a conseguir melhores resultados.

## 6.2 Motivação

A empresa apresentada pretende que os seus agentes de decisão consigam navegar melhor pelo histórico da informação que possui, de forma a conseguir identificar mais rapidamente possíveis problemas ou situações potencialmente críticas para os seus processos de negócio. Para isso a empresa quer focar o seu agente de decisão apenas na informação que lhe diz respeito ou que eventualmente possa estar interessado. A título de exemplo, e uma vez que o 'João Carvalho' é apenas responsável pela parte de vendas de refrigerantes na zona sul do país, as restantes zonas e categorias não são, assim, preferências a aplicar nos seus processos de consulta.

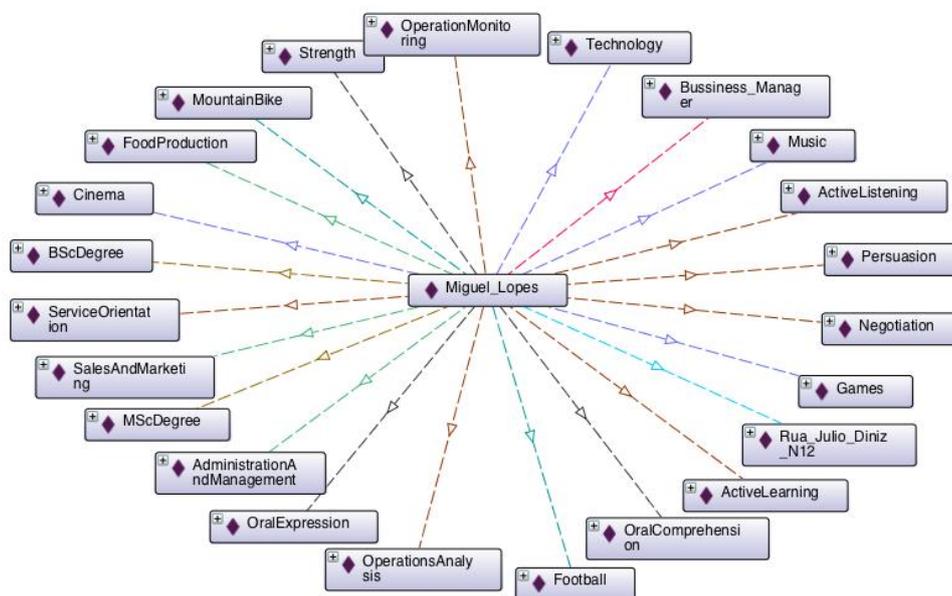


Figura 19 – Caracterização do perfil do 'Miguel Lopes'

## 6.3 Aplicação do Modelo

Nesta secção é apresentado o modelo ontológico das várias personagens, do esquema dimensional utilizado e de algumas consultas e preferências dos utilizadores. Para cada uma das ilustrações incluídas nas secções que se seguem apresentar-se-á um pequeno resumo que caracteriza a personagem envolvida.

### 6.3.1 Perfil do Utilizador

De forma a conseguir ilustrar a aplicação do modelo foi construído um perfil para cada umas das personagens caracterizadas anteriormente.

O 'Miguel Lopes' tem o cargo de 'gestor de negócio' na empresa, sendo responsável pela área de vendas de 'Legumes e Frutos'. Vive na 'Rua Júlio Diniz' em 'Barcelos', é licenciado em 'Gestão' e tem um mestrado em 'Economia Industrial e da Empresa' pela 'Universidade do Minho'. A 'música', o 'cinema' e a 'tecnologia' são alguns dos seus interesses. Também gosta de 'jogar futebol' e de 'jogar computador'. A 'persuasão' e a 'negociação' são alguns dos seus pontos fortes. Tem conhecimentos em 'produção de alimentos', 'vendas' e 'marketing'. A caracterização do seu perfil está ilustrado na Figura 19.

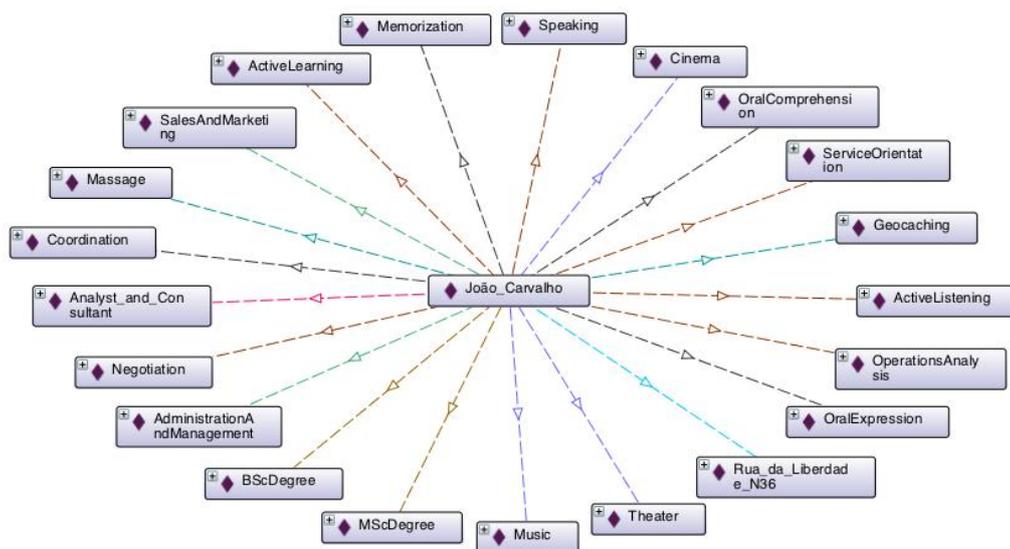


Figura 20 – Caracterização do perfil do 'João Carvalho'

Por sua vez, o 'João Carvalho' tem o cargo de 'gestor de negócio' e é responsável pela área de vendas de 'refrigerantes e vinhos' na 'zona sul' do país. Vive na 'Rua da Liberdade' em 'Matosinhos', é licenciado em 'Administração de Empresas' e tem um mestrado em 'Marketing' pelo 'Instituto Superior do Porto'. Gosta de ir ao 'teatro', 'ouvir música' e não dispensa regularmente uma 'sessão de massagens'. Além disso, tem facilidade na 'compreensão oral' e 'expressão oral', o

que lhe dá vantagens claras em processos de 'negociação'. A caracterização do seu perfil está ilustrado na Figura 20.

Quanto ao 'Pedro Sousa', este tem o cargo de 'gestor de negócio', sendo responsável pela área de vendas de 'refrigerantes e vinhos' na 'zona norte e centro' do país. Vive na 'Rua da Esquina' em 'Guimarães' e é licenciado em 'Negócios Internacionais'. Gosta de 'ouvir música', jogar 'ténis' e 'golfe'. Tem conhecimentos em 'vendas', 'serviço de atendimento a clientes' e 'processos de produção'. O 'desenvolvimento de estratégias de vendas' é um dos seus pontos fortes, possuindo bastante 'originalidade' no desenrolar das suas ações e desenvolvimento dos seus trabalhos. A caracterização do seu perfil está ilustrado na Figura 21.

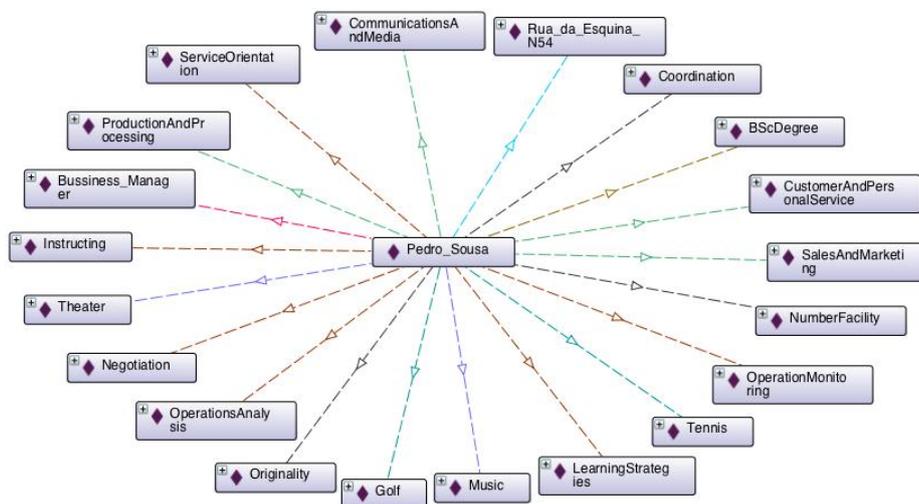


Figura 21 - Caracterização do perfil do 'Pedro Sousa'

Por fim, a 'Raquel Cruz'. Esta tem o cargo de 'gestora de negócio', sendo responsável pela área de vendas de 'congelados'. Vive na 'Avenida da Praia', em 'Setúbal', e tem o '12º ano'. Gosta de 'dançar', 'jogar\_ténis' e fazer 'safaris' com frequência. 'Processos de qualidade' e 'monitorização de operações' são alguns dos seus pontos fortes. Tem facilidade em 'memorizar', 'escrever' e 'compreender'. Como características complementares, tem também grande 'flexibilidade' e 'equilíbrio' nas várias disciplinas de ginástica que esporadicamente realiza. A caracterização do seu perfil está ilustrado na Figura 20.

### 6.3.2 O Esquema Multidimensional

O DW que utilizámos é bastante conhecido na área do retalho (Anexo A). É um esquema típico, organizado em floco de neve, que foi desenvolvido para permitir fazer análises sobre as vendas realizadas, quer seja por produto, categoria de produto, marca, local de armazenamento (cidade, região e país), demografia do cliente, ou, também, por promoções individuais e categorias de promoções. As métricas integradas na tabela de factos do esquema em questão podem ser analisadas pelos diversos atributos da dimensão temporal (dia, semana, mês, trimestre, entre outros).

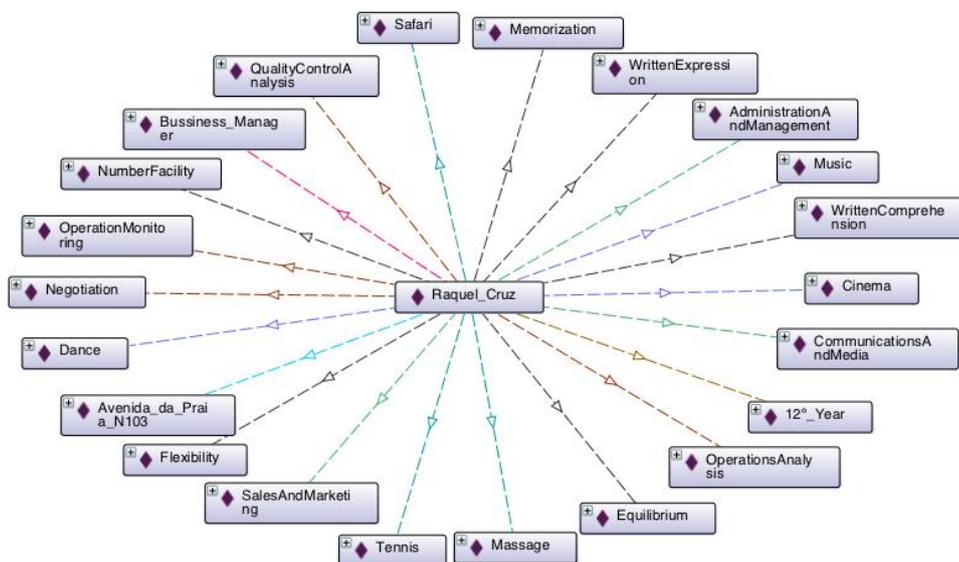


Figura 22 – Caracterização do perfil da 'Raquel Cruz'

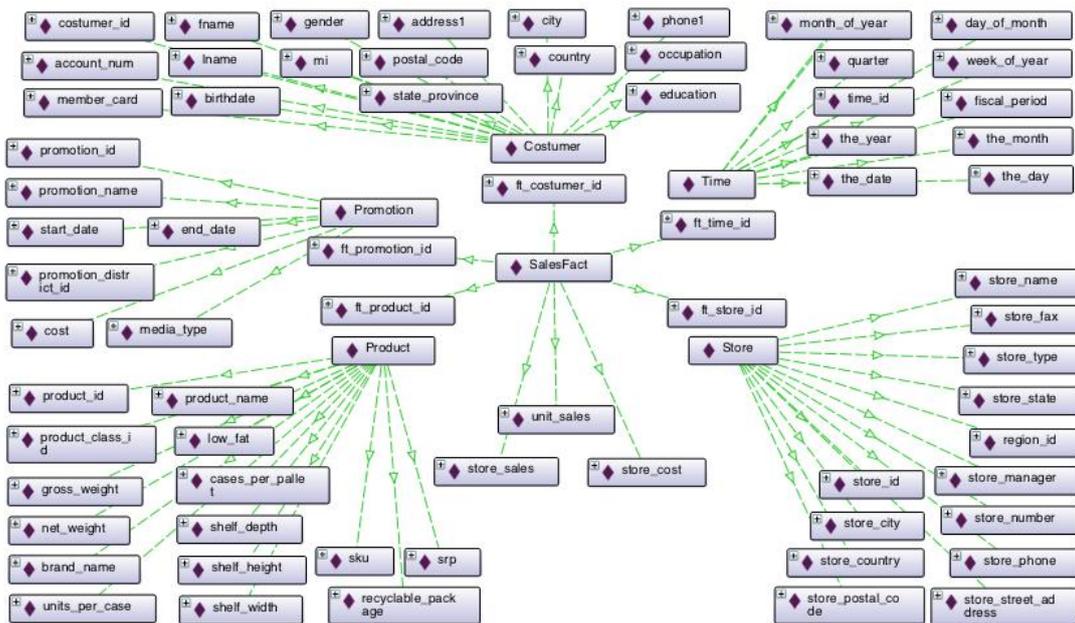


Figura 23 – O esquema multidimensional utilizado

O modelo apresentado na Figura 23 caracteriza a parte do esquema multidimensional que utilizámos. Contudo, foram retirados alguns dos atributos com menor relevância e com o objetivo de simplificar um pouco o esquema e reduzir o número de objetos de dados a manipular.

### 6.3.3 Consultas e Preferências de Utilizador

Como vimos, o 'João Carvalho' é responsável de vendas de refrigerantes e vinhos na zona do sul do país. Durante uma qualquer sessão de exploração OLAP, as suas interrogações, tipicamente, vão se concentrar sobre a categoria de 'refrigerantes e vinhos'. Logo:

- POS(Category, {Refrigerantes, Vinhos})

Visto que o 'João Carvalho' apenas tem à sua responsabilidade a 'zona sul' do país, então:

- POS(StoreRegion, {Lisboa, Setúbal, Évora, Beja, Faro})

Quanto ao 'Pedro Sousa', este está na mesma categoria de produtos do 'João Carvalho', mas na 'zona norte e centro' do país. Isto permite que possamos escrever:

- POS(Category, {Refrigerantes, Vinhos})
- POS(StoreRegion, {Braga, Viana do Castelo, Porto, Aveiro, Coimbra})

Por sua vez, o 'Miguel Lopes' é o responsável de vendas de 'legumes e frutas' em todo o país. Então:

- POS(Category, {Legumes, Frutas})

A 'Raquel Cruz', sendo a responsável pelas vendas de 'congelados', tem todo o interesse em que as suas consultas se "prendam" principalmente nesta categoria, o que faz com que definamos:

- POS(Category, {Refrigerantes, Vinhos})

A simples expressão das preferências acima indicadas, que derivam da categoria em que estão, consegue de imediato focar o utilizador nos produtos que lhe são mais importantes. Assim, cada vez que uma consulta é executada, a aplicação consegue retribuir melhores resultados de uma forma que podemos dizer ser mais "inteligente".

Tal como tem vindo a ser referido ao desta dissertação, uma das principais características das ontologias é a de permitirem a partilha de informação. De forma a conseguir obter esta partilha de informação, é necessário acrescentar ao modelo um contexto. Ou seja, efetuar a criação de grupos de pessoas no qual as suas preferências possam ser partilhadas. Estes grupos devem ser organizados por áreas de interesse, para que a partilha de informação seja mais eficaz. O contexto a ser criado tem que estar ajustado a cada realidade, não podendo ser definido um modelo genérico.

Neste caso de estudo, em particular, o contexto foi a criado através da organização e das suas respectivas áreas de negócio. A Figura 24 ilustra um modelo ontológico (simplista) de uma organização.

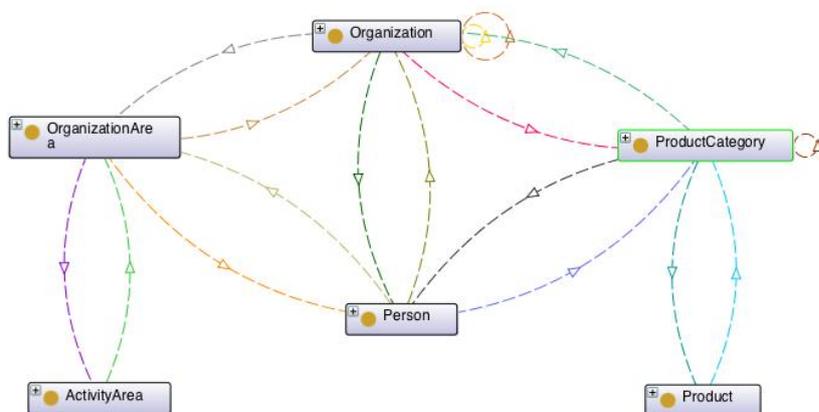


Figura 24 – Modelo de ontologia para uma organização

De uma forma bastante resumida, este modelo introduz 3 novos conceitos importantes, a organização, a categoria/produto e a área de atividade. Assim, torna-se mais fácil a partilha de conhecimento entre os vários agentes de decisão da organização. A Figura 25 ilustra os agentes de decisão no contexto de uma organização.

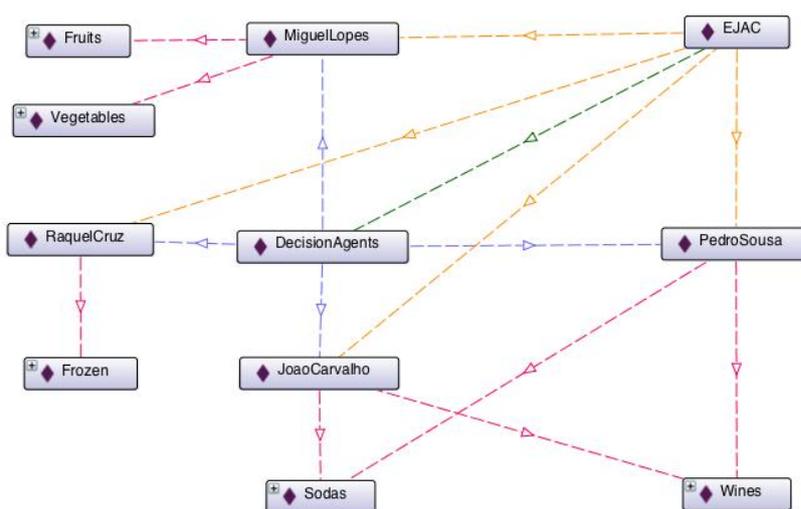


Figura 25 – Caracterização dos agentes de decisão na organização "EJAC"

A título de exemplo, o 'Pedro Sousa' e o 'João Carvalho' trabalham sobre os mesmos produtos, mas são responsáveis por áreas diferentes do país. As suas preferências vão ser facilmente partilháveis. Caso o 'João Carvalho' não tivesse estabelecido a sua preferência pelas categorias de 'refrigerantes e vinhos', quando se efetuasse uma consulta ao sistema, este, através das preferências do 'Pedro Sousa' iria apenas devolver o resultado da consulta sobre as categorias de 'refrigerantes e vinhos'.

## **6.4 Sumário**

O cenário estabelecido neste estudo apresenta exemplos que podem ser derivados de um qualquer domínio da área do retalho. A introdução destes modelos ontológicos não contribuem diretamente para o desenvolvimento dos sistemas de informação, mas são um ponto de partida para o desenvolvimento de abordagens metodicamente bem fundamentadas. Na maioria dos sistemas, as preferências dos utilizadores são estabelecidos através de cláusulas duras (condições "where"), obrigando o utilizador a introduzir um conjunto de preferências, que em muitas situações não conduz a qualquer resultado. O chamado efeito de resultado vazio.



## **Capítulo 7**

### **Conclusões e Trabalho Futuro**

#### **7.1 Um Comentário Final**

Neste trabalho de dissertação foi proposto um modelo ontológico modular, partilhável e genérico, com o objectivo de conseguir representar as preferências dos utilizadores de sistemas OLAP. É descrito, através de princípios e métodos da Web semântica, como é possível melhorar a experiência de utilização dos utilizadores durante uma sessão OLAP. A definição do modelo foi marcada por uma pesquisa bastante detalhada e cuidada de tudo aquilo que outros já tinham realizado anteriormente. A maior parte dos esforços realizados, com vista à melhoria da experiência de utilização dos utilizadores, tem recaído sobretudo no nível da camada de apresentação dos sistemas. Foi, também, objectivo deste trabalho de dissertação mostrar que existem outras direções a seguir nesta problemática.

Os (administradores dos) sistemas de informação necessitam de compreender melhor quem são os seus clientes e o que realmente eles procuram. Esta necessidade surge usualmente com o aumento da informação, provocado na sua maioria pela ligação a serviços de Internet. Numa primeira fase trabalho foi efectuado o estudo dos trabalhos realizados no domínio das preferências de utilização, no qual foram encontradas várias abordagens distintas. As que considerámos mais relevantes foram sendo citadas ao longo desta dissertação. Depois de ter adquirido o conhecimento necessário sobre expressão de preferências, seguiu-se um estudo sobre a Web semântica, nomeadamente em tudo aquilo que se relaciona com ontologias.

As ontologias têm provado que são um meio eficaz para modular conjuntos de informação em formato digital. Tal como mostramos ao longo deste trabalho são uma ferramenta muito útil, pois conseguem apresentar uma visão geral do domínio relacionado a uma área específica. Conseguem ainda um alto nível de abstração, proporcionando também uma semântica rica para que os modelos desenvolvidos possam ser interpretados e processados por máquinas.

## **7.2 O Modelo de preferências**

A problemática das preferências de utilizador tem vindo a ser aplicado em diversas áreas. Analisando a literatura, a noção de preferência foi inicialmente introduzida nos sistemas de base de dados, ainda que apenas a um nível lógico. Posteriormente, seguiu-se a transposição dos conceitos, ainda que um pouco tímida, para os sistemas OLAP e só depois a introdução do conceito da Web semântica no domínio da expressão de preferências. Seguindo as boas prática da Web semântica, fizemos um reaproveitamento de modelos já desenvolvidos.

O modelo proposto neste trabalho assenta principalmente em três componentes distintos, que compõe o modelo de preferências desenvolvido: perfil do utilizador, esquema multidimensional e preferências OLAP. O primeiro componente incorpora uma caracterização quanto aos interesses, atividades, habilidades e capacidades do utilizador. As preferências são sobretudo retiradas através do perfil do utilizador - esta parte foi desenvolvida com base no trabalho realizado por Golemati et. al., 2007). O segundo componente consiste na caracterização de uma base de dados multidimensional, para suportar a definição das preferências do utilizador ao nível das dimensões, tabelas de facto e atributos. Não foi encontrado nenhum trabalho revelante que apresente uma solução de um modelo ontológico para a representação de uma base de dados multidimensional, resulta assim uma primeira abordagem ao problema. Quanto ao terceiro componente, este trata da representação das preferências de utilizadores. O modelo ontológico de preferências apresentado, segue principalmente as abordagens efectuadas por Chomicki (2003), Kießling (2002) e Kießling e Köstler (2002), abordagens estas mais maduras e que apresentaram resultados bastante concretos. Por fim, foi apresentado a ligação entre estes três componentes, resultando assim num único modelo capaz suportar as preferências dos utilizadores em sistemas OLAP. Além da reutilização, o mais importante do modelo obtido é a sua independência quanto ao seu domínio de aplicação ou tecnologia.

### **7.3 Trabalho futuro**

Existe ainda um quinto componente (assumindo que o quarto é o contexto da organização, apresentado no caso de estudo) designado por contexto de ambiente, que especifica o âmbito (domínios) e o meio (dispositivo, localização, etc.) em que determinada preferência foi expressa pelo utilizador. Cada um dos componentes pode ser visto como uma ontologia independente, em que a intersecção destes vários conceitos permite obter uma maior expressividade e uma melhor compreensão em torno das expressão de preferências do utilizador, formando assim uma única ontologia. A implementação deste modelo, na prática, seria o passo seguinte, demonstrando assim o seu real funcionamento. Com este passo, temos a certeza que novos desafios se colocariam. Algo a ser tratado no futuro por uma nova linha de ação de investigação e desenvolvimento no domínio das preferências OLAP.



## Bibliografia

- Sameet Agarwal, Rakesh Agrawal, Prasad M. Deshpande, Ashish Gupta, Jeffrey F. Naughton, Raghu Ramakrishnan e Sunita Sarawagi: "On the Computation of Multidimensional Aggregates". Proceedings of the 22<sup>th</sup> International Conference on Very Large Databases. Mumbai (Bombaim), Índia. Setembro, 1996.
- Agrawal, R. and Wimmers, E. L. 2000. A framework for expressing and combining preferences. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data. SIGMOD '00. ACM, New York, NY, USA, 297–306.
- Börzsönyi, S., Kossmann, D., and Stocker, K. 2001. The skyline operator. In Proceedings of the 17th International Conference on Data Engineering. IEEE Computer Society, Washington, DC, USA, 421–430.
- Chomicki, J. 2003. Preference formulas in relational queries. ACM Trans. Database Syst. 28, 427–466.
- Gaasterland, T. and Lobo, J. 1994. Qualified answers that reflect user needs and preferences. In Proceedings of the 20th International Conference on Very Large Data Bases. VLDB '94. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 309–320.
- Golfarelli, M. and Rizzi, S. 2009. Expressing olap preferences. In Proceedings of the 21st International Conference on Scientific and Statistical Database Management. SSDBM 2009. Springer-Verlag, Berlin, Heidelberg, 83–91.

- Golfarelli, M., Rizzi, S., and Biondi, P. 2010. myolap: An approach to express and evaluate olap preferences. *IEEE Transactions on Knowledge and Data Engineering* 99, PrePrints.
- Kießling, W. 2002. Foundations of preferences in database systems. In *Proceedings of the 28th international conference on Very Large Data Bases. VLDB '02. VLDB Endowment*, 311–322.
- Kießling, W., Hafenrichter, B., Fischer, S., and Holland, S. 2001. Preference xpath: A query language for e-commerce.
- Kießling, W. and Köstler, G. 2002. Preference sql: design, implementation, experiences. In *Proceedings of the 28th international conference on Very Large Data Bases. VLDB '02. VLDB Endowment*, 990–1001.
- Koutrika, G. 2006. Query personalization based on user preference. In *Proceedings of the 20<sup>th</sup> International Conference on Data Engineering. Washington, DC, USA*.
- Koutrika, G. and Ioannidis, Y. 2004. Personalization of queries in database systems. In *Proceedings of the 20th International Conference on Data Engineering. ICDE '04. IEEE Computer Society, Washington, DC, USA*, 597–.
- Lacroix, M. and Lavency, P. 1987. Preferences; putting more knowledge into queries. In *Proceedings of the 13th International Conference on Very Large Data Bases. VLDB '87. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA*, 217–225.
- Mouloudi, H., Bellatreche, L., Giacometti, A., and Marcel, P. 2006. Personalization of mdx queries. In *BDA (2006-09-14)*, D. Laurent, Ed.
- Stefanidis, K., Pitoura, E., and Vassiliadis, P. 2005. On supporting context-aware preferences in relational database systems. In *MCMP*.
- Xin, D. and Han, J. 2008. P-cube: Answering preference queries in multi-dimensional space. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering. IEEE Computer Society, Washington, DC, USA*, 1092–1100.

- Ceccaroni, L. (2001). *Ontoweeds - An ontology-based environmental decision-support system for the management of wastewater treatment plants*. PhD thesis, Universitat Politècnica de Catalunya, Barcelona.
- Guarino, N. 1995. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, 43:625–640.
- Noy, N. F. and McGuinness, D. L. 2003. *Ontology development 101: A guide to creating your first ontology*. Technical report, Stanford University.
- McGuinness, D.L., Fikes, R., Rice, J. and Wilder, S. 2000. An Environment for Merging and Testing Large Ontologies. *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR2000)*. A. G. Cohn, F. Giunchiglia and B. Selman, editors. San Francisco, CA, Morgan Kaufmann Publishers.
- Stevens, R., C.A. Goble, and S. Bechhofer, *Ontology-based knowledge representation for bioinformatics*. *Briefings in Bioinformatics*, 2000. 1(4): p. 398-416.
- A. Farquhar, R. Fikes, and J.P. Rice. The ontolingua server: A tool for collaborative ontology construction. *Journal of Human-Computer Studies*, 46:707–728, 1997.
- Golemati, M. et al., 2007. Creating an Ontology for the User Profile : Method and Applications. *Proceedings of the First RCIS Conference*, p.407–412.
- J. Trajkova, S. Gauch, Improving Ontology-based User Profiles, Proc. of RIAO 2004, University of Avignon (Vaucluse), France, April 26-28, 2004, pp. 380-389
- V. Katifori, A. Poggi, M. Scannapieco, T. Catarci, and Y. Ioannidis (2005). *OntoPIM: how to rely on a personal ontology for Personal Information Management*. In Proc. of the 1st Workshop on The Semantic Desktop.
- S. Lawrence 2000. Context in web search. *IEEE Data Engineering Bulletin*, 23(3):25-32.

- Tapucu, D., Diallo, G., Aït-Ameur, Y. and Ünalir, M.O. 2009. Ontology-Based Database Approach for Handling Preferences, Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction, ISBN: 978-1-60566-756-0, Edited by Ladjel Bellatreche, IGI Global.
- Siberski, W., Pan, J. Z. and Thaden, U. 2006. Querying the semantic web with preferences. In International Semantic Web Conference, 612–624.
- Chamiel, G., Pagnucco, M. 2008: Exploiting ontological information for reasoning with preferences. In: Multidisciplinary Workshop on Advances in Preference Handling.
- Hafenrichter, B., and Kießling, W. 2005. Optimization of Relational Preference Queries. In Proc. of the 16th Australasian Database Conference, 175-184.
- Resnick, P. Varian, H. R. 1997. Recommender systems. *Comm. of the ACM*, 40(3), 56-58, Mar 1997
- Middleton, S.E. Alani, H. Shadbolt, N.R. De Roure, D.C. 2002. Exploiting Synergy Between Ontologies and Recommender Systems. In International Workshop on the Semantic Web, Proceedings of the 11th International World Wide Web Conference WWW-2002, Hawaii, USA.
- Schickel-Zuber, V., and Faltings, B. 2006. Inferring User's Preferences using Ontologies. In *AAAI 2006*, 1413–1418.
- Balabanovic, M., Sholam, Y., 1997. Combining content-based and collaborative recommendation. *Communications of the ACM* 40 (3).
- Kalles, D., Papagelis, A., Zaroliagis, C., 2003. Algorithmic aspects of web intelligent systems. In: Zhong, N., Liu J., Yao, Y. (Eds.), *Web Intelligence*. Springer, Berlin, pp. 323-345.

- Herlocker, J.L., Konstan, J.A., Riedl, J., 2000. Explaining collaborative filtering recommendations. Proceedings of the ACM Conference on Computer Supported Cooperative Work.
- Hofmann, T., 2003. Collaborative filtering via Gaussian probabilistic latent semantic analysis. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.
- Fröschl, C., 2005. User Modeling and User Profiling in Adaptive E-learning Systems. Graz, Austria: Master Thesis.
- Dilek Tapucu, Stephane Jean, Yamine Aït-Ameur and Murat Osman Ünalir, "An Extension of Ontology Based Databases to Handle Preferences", Proceedings of 11th International Conference on Enterprise Information Systems (ICEIS 2009), pp. 208-214, ISBN 978-989-8111-84-5, Milan, Italy, May 2009.
- Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.
- X. Su and T. M. Khoshgoftaar, 2009. A Survey of Collaborative Filtering Techniques, Advances in Artificial Intelligence.
- Ayres, L., Furtado, V., 2007. OWLPref: Uma Representação Declarativa de Preferências para Web Semântica. Anais do XXVII Congresso da SBC.
- Bontas, P. E., Mochol, M. and Tolksdorf, R. 2005. Case Studies on Ontology Reuse. In Proc. of the 5th International Conference on Knowledge Management.
- Golemati M., Katifori A., Vassilakis C., Lepouras G., Halatsis C. 2007. Creating an Ontology for the User Profile: Method and Applications, in Proc. Of the First RCIS Conference, Ouarzazate, Morocc.
- Ngoc, K.A.P., et al, 2005, OWL-Based User Preference and Behavior Routine Ontology for Ubiquitous System. Lecture notes in computer science, Vol. 3761, pp. 1615-1622.
- Kunder, M., 2007. *The size of World Web Web (The Internet)*. [online] Available at:

---

<http://www.worldwidewebsize.com> [Accessed 14 September 2012].

Gruber, T., 2000. *Sharable Ontologies Library*. [online] Available at: <http://ksl-web.stanford.edu/knowledge-sharing/ontologies/> [Accessed 7 September 2012].

Gruber, T., 2009. *Encyclopedia fo Database Systems*. [online] Springer-Verlag. Available at: <http://tomgruber.org/writing/ontology-definition-2007.htm> [Accessed 3 October 2012].

Loom, 2007. *Loom*. [online] Available at: <http://www.isi.edu/isd/LOOM/> [Accessed 21 October 2012].

KMI, 2004. *Operational Conceptual Modelling Language*. [online] Available at: <http://technologies.kmi.open.ac.uk/ocml/> [Accessed 22 October 2012].

W3C, 2007. *Semantic Web Activity*. [online] Available at: <http://www.w3.org/sw/> [Accessed 12 July 2012].

Obitko, M., 2007. *Introduction to Ontologies and Semantic Web*. [online] Available at: <http://www.obitko.com/tutorials/ontologies-semantic-web/> [Accessed 12 July 2012].

W3C, 2007. *SPARQL Query Language for RDF*. [online] Available at: <http://www.w3.org/TR/rdf-sparql-query/> [Accessed 28 July 2012].

Stanford University, 1997. *The Protégé Ontology Editor and Knowledge Acquisition System*. [online] Available at: <http://protege.stanford.edu> [Accessed 18 June 2012].

CIF, 2007. Corporate Information Factory. [online] Available at: <http://www.inmoncif.com/about/> [Accessed 6 December 2012]

Stanford University, 1997. *Ontolingua*. [online] Available at: <http://www.ksl.stanford.edu/software/ontolingua/> [Accessed 28 July 2012].

Flora, 2008. *Frame Logic*. [online] Available at: <http://flora.sourceforge.net/aboutFlogic.php> [Accessed 29 July 2012].

- 
- Karp, R., Chaudhri, V. and Thomere, J. ,1999. *XOL: An XML-based ontology exchange language (version 0.4)*. [online] Available at: <http://www.ai.sri.com/~pkarp/xol> [Accessed 29 July 2012].
- Beckett, D., 2004. *RDF/XML Syntax Specification, W3C Recommendation*. [online] Available at: <http://www.w3.org/TR/rdf-syntax-grammar/> [Accessed 29 July 2012].
- Brickley, D. and Guha, R. V., 2004. *RDF vocabulary description language 1.0: RDF Schema. W3C Recommendation*. [online] Available at: <http://www.w3.org/TR/rdf-schema/> [Accessed 29 July 2012].
- OIL, 2000. *Ontology Interchange Language*. [online] Available at: <http://www.ontoknowledge.org/oil/index.shtml> [Accessed 30 July 2012].
- Heflin, J., Hendler, J. and Luke, S., 1999. SHOE: A knowledge representation language for internet applications. Technical Report CS-TR-4078, College Park: University of Maryland.
- McGuinness, D. L., and Harmelen, F. V., 2004. *OWL web ontology language overview. W3C Recommendation*. [online] Available at: <http://www.w3.org/TR/owl-features/> [Accessed 12 April 2012].
- DAML, 2006a. *The DARPA Agent Markup Language*. [online] Available at: <http://www.daml.org/> [Accessed 28 July 2012].
- DAML, 2006b. *DAML Ontology Library*. [online] Available at: <http://www.daml.org/ontologies/> [Accessed 28 July 2012].
- GO, 1999. *The Gene Ontology*. [online] Available at: <http://www.geneontology.org> [Accessed 4 August 2012].
- Prud'hommeaux, E. and Seaborne, A., 2006. SPARQL Query Language for RDF. W3C Working Draft.



# Anexo A

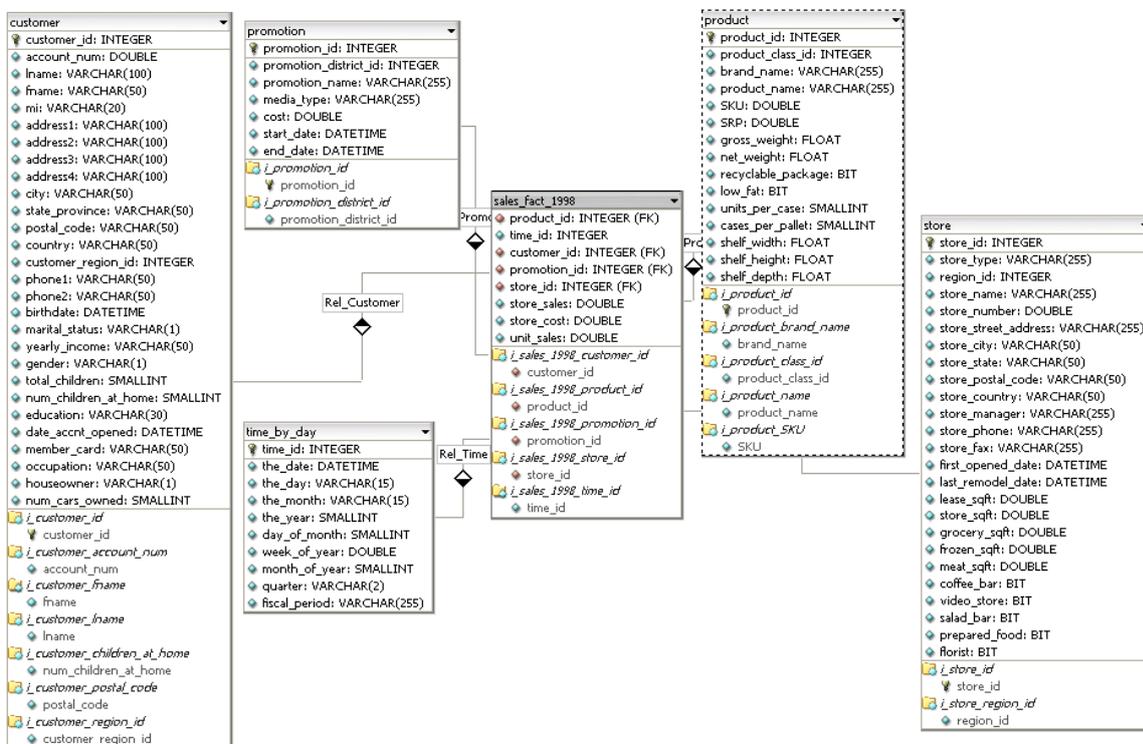


Figura 26 – O esquema multidimensional da "FoodMart"