

Modelagem Da Informação Com Idef1x

Os sistemas de informações estão presentes, cada vez mais, na vida de organizações e indivíduos. A qualidade destes sistemas depende dos sistemas de bancos de dados que os suportam. O projeto de um banco de dados de alta qualidade depende, por sua vez, da modelagem da informação, que trata da concepção da estrutura da informação em determinado universo de discurso (ambiente, empresa, negócio, etc.).

Reingruber & Gregory (1994) sustentam que a importância da modelagem da informação tem sido subestimada. Desenvolvedores de sistemas dedicam, tradicionalmente, maior atenção à modelagem dos processos que usam informações do que à estrutura subjacente.

Esta mentalidade vem mudando à medida em que os sistemas de informações crescem em tamanho e complexidade, evidenciando os problemas causados por modelagem da informação deficiente. Várias abordagens têm sido propostas para a modelagem da informação. Todas, de forma geral, têm base no modelo Entidade-Relacionamento de Peter Chen (1976).

Dentre as mais difundidas encontram-se NIAM (Nijssen & Halpin 1989), método CASE de Barker (1990), Engenharia da Informação (Martin & McClure 1985) e IDEF1X (NIST 1993). As abordagens citadas têm sido bastante utilizadas no projeto de bancos de dados relacionais.

Mais recentemente, técnicas baseadas na orientação a objeto têm sido propostas para a modelagem da informação para projeto de bancos de dados orientados a objeto: UML (Muller 1999) e IDEFobject (IEEE 1998), uma nova versão de IDEF1X.

Os bancos de dados relacionais têm um embasamento teórico sólido (Codd 1970), testado e aprovado durante quase três décadas.

Embora a orientação a objetos seja uma tecnologia mais moderna, a maior parte dos sistemas de informações em desenvolvimento ainda é baseada em bancos de dados relacionais, garantindo uma sobrevida longa a técnicas de modelagem da informação voltadas ao mundo relacional, como IDEF1X.

O PADRÃO IDEF1X IDEF1X é uma linguagem e método para modelagem da informação publicada como padrão norte-americano pelo NIST (1993). A linguagem já era usada, entretanto, antes da existência do padrão (Bruce 1992).

As seções a seguir apresentam as construções da linguagem: entidades, relacionamentos, chaves, atributos, e notas. As cinco fases do método IDEF1X para modelagem da informação são sumariamente descritas. A linguagem IDEF1X As construções sintáticas básicas da linguagem IDEF1X são: entidade, relacionamento e atributo.

Entidades, representadas graficamente como retângulos com uma linha divisória horizontal, são conjuntos de objetos de mesma natureza. Na figura 1, a entidade Produto representa um conjunto de instâncias ou ocorrências de produtos. Relacionamentos são associações significativas entre duas ocorrências de entidades.

São representados por linhas rotuladas com um verbo ou frase verbal, como em Nota Fiscal tem Item. Atributos são características das entidades, como ilustra a figura 1, onde endereço é um atributo de Cliente.

As entidades IDEF1X podem ser dependentes, representadas por retângulos com os cantos arredondados, ou independentes, representadas por retângulos com os cantos vivos. Uma entidade é dependente sempre que sua chave (também conhecida como chave primária) contém a chave completa de pelo menos uma outra entidade.

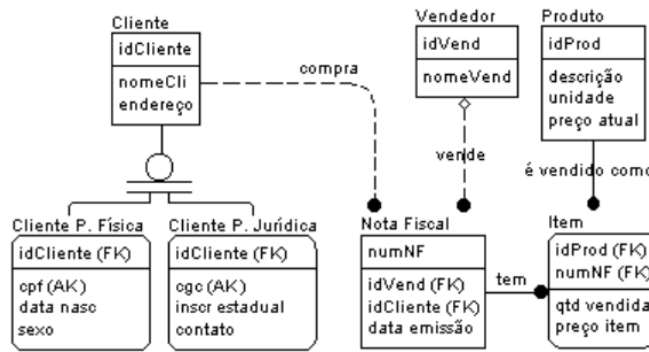


Figura 1 – Modelo IDEF1X ilustrando as construções básicas da linguagem

A chave de uma entidade é formada por um ou mais atributos cujo valor identifica uma ocorrência da entidade. Os atributos chaves são inscritos acima da divisão horizontal do retângulo que representa a entidade. Além da chave primária, IDEF1X descreve outros dois tipos de chave: estrangeira e alternada. Uma chave estrangeira é anotada pelo símbolo (FK) (foreign key), o qual denota a origem estrangeira do atributo, que é migrado a partir de outra entidade, onde é chave primária.

Na figura 1, idCliente é um atributo chave estrangeira em Nota Fiscal, migrado desde Cliente. Chave alternada é o atributo (ou conjunto de atributos) cujo valor identifica uma ocorrência da entidade, mas não foi escolhido como chave. Um atributo chave alternada é anotado pelo símbolo (AK) (alternate key), como no exemplo da figura 1, onde cgc é um identificador de Cliente P. Jurídica.

A migração do atributo chave de uma entidade e sua implementação como chave estrangeira em outra entidade respeita a regra de migração de chaves: todo relacionamento implica na migração da chave completa, e toda migração de chave é devida à existência de relacionamento. A forma como a migração ocorre varia segundo o tipo de relacionamento.

A figura 2 apresenta esquematicamente os tipos de relacionamento IDEF1X e suas representações gráficas. Relacionamentos podem ser específicos ou não-específicos.

Um relacionamento específico é aquele que, do ponto de vista de pelo menos uma das ocorrências de entidade envolvidas, tem uma outra ocorrência específica associada segundo o relacionamento. Relacionamentos não-específicos, por outro lado, são associações várias-paravárias, nas quais não é possível especificar uma (pois são várias) ocorrência associada.

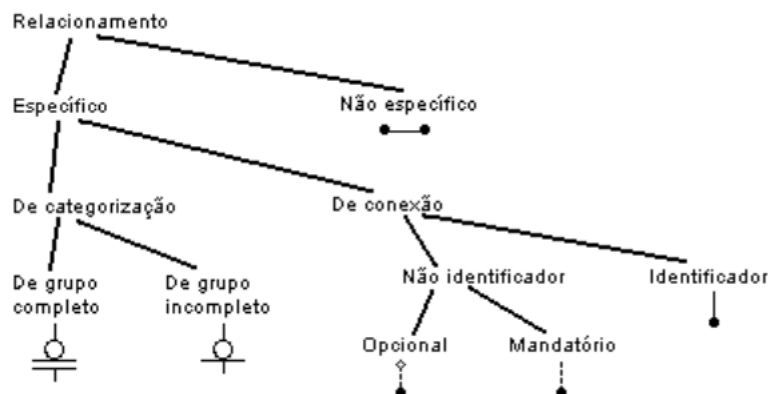


Figura 2 - Tipos de relacionamento na linguagem IDEF1X

Os relacionamentos específicos podem ser de categorização ou de conexão. Em um relacionamento de categorização, ocorrências de ambas entidades associadas representam o mesmo tipo de objeto real ou abstrato, sendo que a entidade genérica representa os atributos genéricos do objeto, enquanto cada entidade categoria reúne características de uma determinada especialização.

Nos relacionamentos de conexão, também chamados de conexão específica, qualquer número de ocorrências de uma entidade filha (child entity) associa-se a zero ou uma ocorrência de uma entidade genitora (parent entity).

Vários relacionamentos de categorização podem reunir entidades categorias em um grupo de categorização (category cluster), representado por uma circunferência sublinhada por um ou dois traços. Uma ocorrência da entidade genérica pode estar associada a apenas uma ocorrência de uma das entidades categorias do grupo (isto é, as categorias são excludentes).

Um grupo completo é anotado por uma linha dupla sob a circunferência que representa o grupo. Todas as categorias estão presentes, ou seja, uma ocorrência da entidade genérica está necessariamente associada a uma das categorias, como no caso da entidade genérica Cliente, na figura 1, onde qualquer ocorrência está associada a uma ocorrência da entidade categoria Cliente P.

Física ou (exclusivamente) Cliente P. Jurídica. Em um grupo incompleto, nem todas as categorias possíveis estão presentes. Uma ocorrência da entidade genérica pode ou não estar associada a uma das categorias. Os relacionamentos de conexão específica nos quais toda a chave da entidade genitora faz parte da chave da entidade filha são ditos identificadores.

O relacionamento Nota Fiscal tem Item, na figura 1, é um exemplo de relacionamento identificador, pois a chave de Nota Fiscal faz parte da chave de Item. Relacionamentos de conexão onde pelo menos um atributo chave da entidade genitora não faz parte da chave da entidade filha chamam-se não-identificadores, e são representados por linha pontilhada.

Podem ser mandatórios, quando toda ocorrência da entidade filha deve estar associada a uma ocorrência da entidade genitora, ou opcionais (representados por um pequeno losango junto à entidade genitora), nos casos nos quais é possível que uma ocorrência da entidade filha não esteja associada a uma ocorrência da entidade genitora.

Na figura 1, Cliente compra Nota Fiscal é um relacionamento mandatório, enquanto Vendedor vende Nota Fiscal é opcional – permite que haja nota fiscal sem vendedor associado. IDEF1X permite representar precisamente a cardinalidade dos relacionamentos, que diz respeito a quantas ocorrências podem estar associadas a uma ocorrência de entidade envolvida em determinado relacionamento.

Todo relacionamento deve ser considerado nos dois sentidos possíveis: da entidade genitora (ou genérica) para a filha (ou categoria), e vice-versa. No sentido da entidade filha (ou categoria) para a genitora (ou genérica), a cardinalidade é “exatamente uma”, com a exceção dos relacionamentos não-identificadores opcionais, onde a cardinalidade é “zero ou uma”.

No sentido inverso, a variedade é maior. A figura 3 apresenta a representação gráfica das várias cardinalidades possíveis para relacionamentos de conexão e seus significados. Em relacionamentos de categorização, a cardinalidade no sentido da entidade genérica para a categoria é sempre “zero ou uma”.

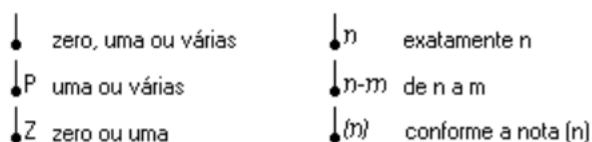


Figura 3 - Cardinalidade de relacionamentos de conexão

Restrições sobre a estrutura da informação que não são representáveis através dos elementos da linguagem recém descritos são representados através de notas.

Um exemplo é ilustrado na figura 3, onde um índice (n) junto à entidade filha referencia uma nota (que deve ser apresentada em português, indexada, em outro lugar no diagrama) que descreve uma cardinalidade não convencional. A notação de IDEF1X permite construir modelos precisos e detalhados.

Às qualidades da linguagem unem-se as virtudes do método IDEF1X, que reúne a experiência de muitos anos da indústria e forças armadas norte-americanas. O método IDEF1X é apresentado a seguir.

O Método IDEF1X

O padrão IDEF1X (NIST 1993) apresenta um guia de projeto para o desenvolvimento de um modelo da informação.

Este guia, ou método IDEF1X, propõe o desenvolvimento de um modelo da informação em cinco fases, com detalhamento iterativo e progressivo do modelo:

- Fase 0 – Iniciação de projeto.

- Fase 1 – Definição de entidades.
- Fase 2 – Definição de relacionamentos.
- Fase 3 – Definição de chaves.

Fase 4 – Definição de atributos. Na Fase 0 estabelecem-se os objetivos da modelagem (documentados na forma de uma declaração de propósito e escopo), desenvolve-se um plano (incluindo cronograma) de modelagem, organiza-se a equipe de projeto e coleta-se o material a ser usado como fonte de informação.

O material-fonte deve ser indexado e arquivado para futura referência. As figuras 4 e 5 representam o material-fonte coletado em um projeto-exemplo a ser ilustrado nesta seção: a modelagem da informação referente a notas fiscais de venda da Empresa Comercial Fictícia (ECF).

A figura 4 representa o material número 1 (nota fiscal de venda de mercadorias); a figura 5 ilustra o material número 2 (ficha de produto). A figura 6 apresenta os papéis existentes numa equipe de projeto IDEF1X: o gerente de projeto administra o projeto, contrata o modelador, identifica os peritos e fontes de informação adequadas, e preside o comitê de revisão de aceitação do modelo.

O modelador é responsável pela coleta de informações e concepção do modelo da informação, bem como pelo treinamento de revisores e documentação do modelo.

Fontes do modelo são pessoas e documentos que podem oferecer esclarecimento sobre o ambiente modelado. Peritos são pessoas cujo conhecimento e experiência são usados na validação do modelo concebido. O comitê de revisão de aceitação dá a palavra final sobre a qualidade e aceitabilidade do modelo. Para evitar conflito de interesses, o comitê não deve ter a participação de modeladores.

EMPRESA COMERCIAL FICTÍCIA (ECF)			
Rodovia BR 101 s/n - São José			
DATA:	31/08/99	Nº	123456
DESTINATÁRIO/REMETENTE			
Loja Sertão Ltda.			
CGC			
88122000/0001-11			
ENDEREÇO			
Rodovia SC 407 Km 4 São José			
DADOS DO PRODUTO			
CÓDIGO	DESCRIÇÃO	QTDE	UN PREÇO
1233	POLTR GIRAT X	2,00	un 98,00
1344	VENT TETO VT3	1,00	un 38,95
TOTAL			234,95

Figura 4 – Nota fiscal, material-fonte nº 1

Código:	1344
Descrição:	VENT TETO VT3
Unidade:	un
Preço:	41,00
Qtd estoque:	22 24 16

Figura 5 – Ficha de produto, material-fonte nº 2

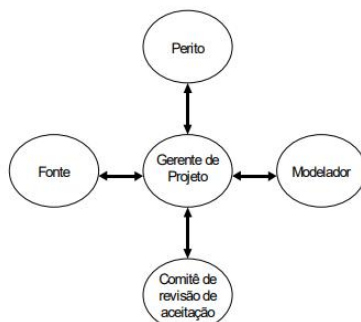


Figura 6 – Papéis numa equipe de projeto IDEF1X (NIST 1993)

As fases 1 e 2 tratam da estrutura da informação num nível de abstração alto. Na Fase 1, as entidades são identificadas e definidas.

Na Fase 2 as associações relevantes entre as entidades concebidas na fase anterior são identificadas e definidas, e um modelo gráfico de entidades e relacionamentos é desenhado.

As definições formam um glossário – são sentenças em português que definem o que é cada ocorrência de cada entidade ou relacionamento. A figura 7 ilustra, na forma de um conjunto de documentos conhecido como kit IDEF1X, os resultados das fases 0, 1 e 2 do projeto-exemplo.

Os documentos apresentados são: declaração de propósito e escopo, organização da equipe de projeto e lista do material-fonte (da Fase 0), lista e glossário de entidades (da Fase 1), e glossário de relacionamentos e diagrama de entidades e relacionamentos (da Fase 2).

O diagrama da Fase 2, ilustrado em separado na figura 8, representa apenas entidades e relacionamentos, sem especificar o tipo, uma vez que este detalhamento depende da especificação das chaves, assunto da Fase 3. A notação de relacionamento identificador é usada indistintamente. Relacionamentos não-específicos podem ser usados na Fase 2, como é o caso do existente entre as entidades Nota Fiscal e Produto, na figura 8.

Projeto: ECF/Nota Fiscal		Data: 1999-08-20
KIT de projeto, fases 0 a 2 (resumido)		
Autor: Onofre Estuff		
<u>Declaração de propósito e escopo:</u>		
O propósito deste modelo é representar a estrutura atual da informação relativa a notas fiscais emitidas pela ECF. As informações de interesse são as relativas à emissão de notas fiscais de venda.		
<u>Equipe de projeto:</u>		
Gerente: Ana Lista (gerente de informática)		
Modelador: Onofre Estuff (responsável)		
Fonte: João Balcão (balconista), manuais da Receita Federal relativos a notas fiscais		
Perito: Jacob Brando (contador)		
Comitê de revisão de aceitação: Ana Lista, Rui D.Querer (consultor), Jacob Brando		
<u>Material-fonte:</u>		
1 - Nota Fiscal de venda de mercadorias		
2 - Ficha de produto		
<u>Relatório de entidades:</u>		
No	Nome da entidade	No.doc.-fonte
E1	Produto	2
E2	Nota Fiscal	1
E3	Cliente	1
<u>Glossário:</u>		
<u>Entidades</u>		
Nome: Produto		
Definição: Produto que pode ser vendido pela ECF.		
Nome: Nota Fiscal		
Definição: Nota fiscal emitida pela ECF.		
Sinônimos: NF, Nota		
Nome: Cliente		
Definição: Cliente cadastrado na ECF.		
<u>Relacionamentos</u>		
Produto é vendido em zero, uma ou várias notas fiscais.		
Nota fiscal tem um ou vários produtos.		
Cliente recebe zero, uma ou várias notas fiscais.		
Nota fiscal é faturada contra um cliente.		
<Diagrama de nível de entidade - figura 8>		

Figura 7 - Kit IDEF1X para o projeto-exemplo (fases 0 a 2)

Em um diagrama IDEF1X, cada relacionamento tem registrados seu e cardinalidade. O nome esclarece qual o caráter do relacionamento entre as ocorrências de entidade envolvidas e é representado por uma frase verbal anotada junto ao relacionamento. Nos casos de relacionamentos específicos, pode-se representar apenas a frase verbal no sentido da entidade genitora (lado específico do relacionamento) para a entidade filha. Na figura 8, por exemplo, representa-se a frase verbal recebe, do relacionamento Cliente – Nota Fiscal, e omite-se a frase verbal é faturada contra um, do relacionamento Nota Fiscal – Cliente.

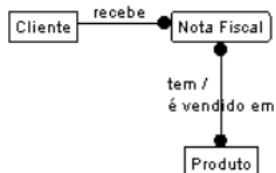


Figura 8 - Diagrama de entidades e relacionamentos do projeto-exemplo

Na Fase 3, os atributos chaves de cada entidade são identificados e definidos. Os tipos dos relacionamentos são detalhados, a migração de chaves e a resolução de relacionamentos não-específicos são executadas.

A resolução de relacionamentos não específicos gera novas entidades e relacionamentos, o que demanda uma atualização do projeto nas fases 1 e 2. Um glossário de atributos chaves é anexado à documentação.

Na Fase 4, os atributos não-chaves são identificados e definidos, e as definições são anexadas à documentação. Cada atributo deve ser único no modelo, isto é, deve ter nome único e ser próprio de apenas uma entidade. A figura 9 apresenta o detalhamento do projeto-exemplo iniciado nesta seção. Neste kit IDEF1X da Fase 4 há uma nova entidade, ItemNF, com dois novos relacionamentos específicos (com Nota Fiscal e Produto), derivados do relacionamento não-específico entre Nota Fiscal e Produto.

A figura 10 apresenta o diagrama completo para o projeto-exemplo, que faz parte do kit da Fase 4. Os kits são usados pelo comitê de revisão de aceitação para verificar e validar o modelo. Opcionalmente, ou adicionalmente, é conduzida uma revisão final, conhecida como procedimento walkthrough.

O modelo recebe um status que determina sua aceitação plena ou condicional, sua reconstrução, ou mesmo sua rejeição. A razão de se determinar a validade do modelo a partir da avaliação de uma equipe de peritos é que não existe um modelo absolutamente “correto”.

Uma concepção de modelagem pode ser considerada adequada quando atende ao que o comitê de revisão considera correto ou satisfatório. Um modelo da informação é desenvolvido para atender a um propósito (registrado na declaração de propósito e escopo da Fase 0), e não para representar “a Verdade”. Esta noção é explorada em maior profundidade na próxima seção.

Projeto: ECF/Nota Fiscal	Data: 1999-10-01
KIT de projeto, fases 3 e 4 (resumido)	
Autor: Onofre Estuff	
<u>Glossário:</u>	
<u>Entidades e atributos</u>	
Nome: Produto	
Definição: Produto que pode ser vendido pela ECF.	
Atributo id Produto: Código identificador de produto	
Atributo descr: Descrição ou nome de produto	
Atributo unidade: Unidade na qual é vendido o produto	
Atributo qtde estoque: Quantidade em estoque do produto	
Atributo preço atual: Preço atual do produto	
Nome: Nota Fiscal	
Definição: Nota fiscal emitida pela ECF.	
Sinônimos: NF, Nota	
Atributo num NF: Número identificador da nota	
Atributo id Cliente: Código do cliente destinatário da nota	
Atributo data emissão: Data da emissão da nota	
Nome: Cliente	
Definição: Cliente cadastrado na ECF.	
Atributo id Cliente: Código identificador do cliente	
Atributo nome Cliente: Nome do cliente	
Atributo end Cliente: Endereço do cliente	
Atributo cgc Cliente: Código identificador do cliente no CGC	
Nome: ItemNF	
Definição: Item de nota; produto vendido em uma nota.	
Atributo num NF: Número da nota à qual se refere o item	
Atributo id Produto: Código do produto vendido no item	
Atributo qtde vendida: Quantidade vendida do produto no item	
Atributo preço venda: Preço praticado na venda do produto neste item	
<u>Relacionamentos</u>	
<Legíveis no diagrama...>	
<Diagrama completo - figura 10>	

Figura 9 - Kit IDEF1X para um projeto-exemplo (completo)

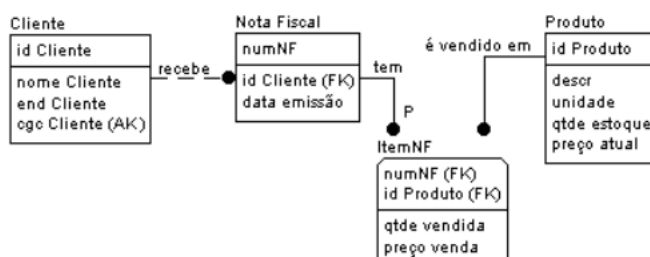


Figura 10 - Diagrama completo do projeto-exemplo

O Melhor De Idef1x: O Princípio Do Consenso

Hay (1995) compara várias linguagens de modelagem, incluindo IDEF1X, quanto à completude técnica e legibilidade. Conclui declarando que não há a linguagem “correta”, e que o sucesso de um projeto de banco de dados depende mais da habilidade do projetista do que da notação escolhida.

A habilidade do modelador é, com efeito, mais importante do que a escolha da linguagem.

Mesmo assim, os estudos comparativos entre as diversas linguagens para modelagem da informação têm mérito, já que detalhes da sintaxe e semântica de cada linguagem podem favorecer ou prejudicar a qualidade de um modelo da informação. Entretanto, há um outro fator crítico para o sucesso de um projeto de banco de dados: a forma como é conduzida a avaliação e aprovação de um modelo.

IDEF1X, ao contrário das demais abordagens, define um método detalhado para ser usado em conjunto com a linguagem. O método é top-down, iterativo, progressivo, e tem como critério para a aprovação de um modelo o consenso de peritos na área de domínio, ou universo de discurso, do modelo. Um modelo não pode ser mais “correto” do que o consenso dos peritos.

Caso contrário, existiria uma verdade absoluta cuja modelagem seria de responsabilidade do modelador. Portanto, o resultado que se pode alcançar em um esforço de modelagem é a captura da estrutura da informação que representa as regras do negócio “corretas” segundo a avaliação dos peritos. Este princípio do consenso foi incorporado a IDEF1X a partir da experiência acumulada pela indústria e forças armadas norte-americanas em projetos de bancos de dados, muitos dos quais de grande tamanho, complexidade, e importância econômica e estratégica.

Ainda assim, nada assegura que o consenso de peritos na avaliação de um modelo representa garantia permanente de qualidade. É possível que mudem as regras do negócio, ou o consenso dos peritos. Portanto, é importante criar condições para a manutenção dos modelos. O glossário com as definições de entidades, relacionamentos e atributos, juntamente com a lista indexada de material-fonte e as declarações de propósito e escopo, são os documentos que permitem a qualquer modelador, participante ou não do esforço de modelagem original, compreender e manter um modelo.

O caráter relativo do consenso de peritos sobre a concepção de modelos de informação pode ser observado em outras modalidades de modelagem, não apenas de bancos de dados. Por exemplo, pode-se observar o mapa-múndi da figura 11 e fazer a pergunta de modelagem: “Quais são os continentes existentes?”

O dicionário Aurélio (Ferreira 1986) apresenta várias definições de continente. As acepções que nos interessam são: “continente. [...] 4. Grande massa de terra cercada pelas águas oceânicas. 5. Cada uma das grandes divisões da terra: Europa, Ásia, África, América e Oceania. 6. O corpo principal da Europa em relação às Ilhas Britânicas. 7. Bras., RS. Designação dada popularmente ao Rio Grande do Sul, desde os tempos coloniais até a revolução de 1835.”

Ao contrário de um glossário IDEF1X, um dicionário permite definições ambíguas e até conflitantes. A acepção 4 é genérica e, certamente, a menos tendenciosa. A acepção 5 impõe a existência de um “continente” europeu e ignora a existência da Antártida, que é maior do que a Europa e a Oceania. As acepções 6 e 7 são muito particulares, válidas em contextos específicos. De acordo com a acepção 4, a Europa não é um continente. Embora as definições 4 e 5 sejam contraditórias, costumam ser ensinadas nos bancos escolares como se fossem compatíveis, sem grandes protestos de geógrafos ou educadores.

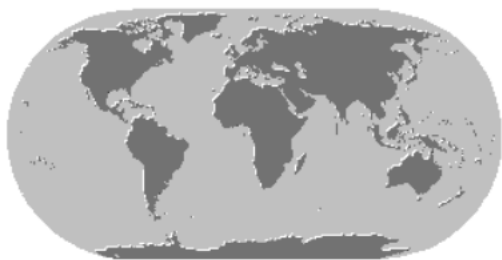


Figura 11 – “Quais são os continentes?”

