deti universidade de aveiro
departamento de electrónica,
telecomunicações e informática

Semantic Web

Engenharia de Dados e Conhecimento

2019/2020

Semântica dos Dados

Semântica (i)

- A Linguagem Natural
  - O melhor “API” para acesso ao Conhecimento.
  - Exemplo de 2 frases simples:
    - “O João gosta de aviões.”
    - “Os aviões assustam a Joana.”
  - Análise às frases:
    - Cada uma das frases é um “pedaço” de informação
    - As palavras “João” e “Joana” referem pessoas específicas
    - A palavra “aviões” refere uma classe de veículos
    - As palavras “gosta” e “assustam” expõe uma relação entre a pessoa e o veículo

Semântica (ii)

- Análise (cont.):
  - existe um conhecimento prévio do significado das palavras “avião”, “gostar” e “assustar”
  - o que permite entender perfeitamente o significado das 2 frases
  - e o que leva à criação de mais conhecimento.
- Daqui sobressai que:
  - símbolos (palavras) referem coisas ou conceitos
  - sequências de símbolos expressam significado ou semântica

- Este é um exemplo de semântica

Conhecimento

- Munido do conhecimento exposto pelas 2 frases anteriores, é possível responder a perguntas, como:
  - Alguém gosta de aviões?
  - Alguém tem medo de aviões?
  - Quem gosta de aviões?
  - Quem tem medo de aviões?
  - É possível as pessoas gostarem de aviões?
  - É possível as pessoas terem medo de aviões?
  - Os aviões são passíveis de serem apreciados (gostar) por pessoas?
  - Os aviões são passíveis de assustar pessoas?

Dados na Web (i)

- A web tornou-se o denominador comum para a interface de aplicações distribuídas
- Esta padronização de facto, levou a que deixasse de ser importante a comunicação da semântica dos dados, mas apenas como estes seriam mostrados
- Esta abordagem tem a vantagem de ter facilitado a explosão de aplicações disponíveis
- Mas tem a desvantagem de esconder os dados e o seu significado, o que dificulta, e até mesmo impossibilita, a sua integração noutras aplicações

Dados na Web (ii)

- Desta forma:
  - aplicações que combinam dados de novas formas
  - e permitem aos utilizadores estabelecer ligações e perceber relações, previamente escondidas
  - são tidas como muito poderosas
  - e a sua existência é “obrigatória” nos atuais contextos de utilização da informação
- Algumas aplicações já conseguem:
  - fazer a recolha e a integração de dados provindos de diversas fontes
  - contudo o seu atual processo de desenvolvimento é altamente especializado e cheio de idiossincrasias

Métodos Tradicionais de Modelação de Dados

- Modelo Tabular
  - Uma forma simples de modular dados
  - Muito familiar à maioria dos utilizadores
  - Exs:
    - dados numa folha de cálculo
    - dados numa tabela HTML
  - A sua grande vantagem é a legibilidade por parte dos utilizadores humanos.

Modelo Tabular (i)

- Exemplo:

Restaurant	Address	Cuisine	Price	Open
Deli Llama	Peachtree Rd	Deli	$	Mon, Tue, Wed, Thu, Fri
Peking Inn	Lake St	Chinese	$$$	Thu, Fri, Sat
Thai Tanic	Branch Dr	Thai	$$	Tue, Wed, Thu, Fri, Sat, Sun
Lord of the Fries	Flower Ave	Fast Food	$$	Tue, Wed, Thu, Fri, Sat, Sun
Marquis de Salade	Main St	French	$$$	Thu, Fri, Sat
Wok This Way	Second St	Chinese	$	Mon, Tue, Wed, Thu, Fri, Sat, Sun
Luna Sea	Autumn Dr	Seafood	$$$	Tue, Thu, Fri, Sat
Pita Pan	Thunder Rd	Middle Eastern	$$	Mon, Tue, Wed, Thu, Fri, Sat, Sun
Award Weiners	Dorfold Mews	Fast Food	$	Mon, Tue, Wed, Thu, Fri, Sat
Lettuce Eat	Rustic Parkway	Deli	$$	Mon, Tue, Wed, Thu, Fri

Modelo Tabular (ii)

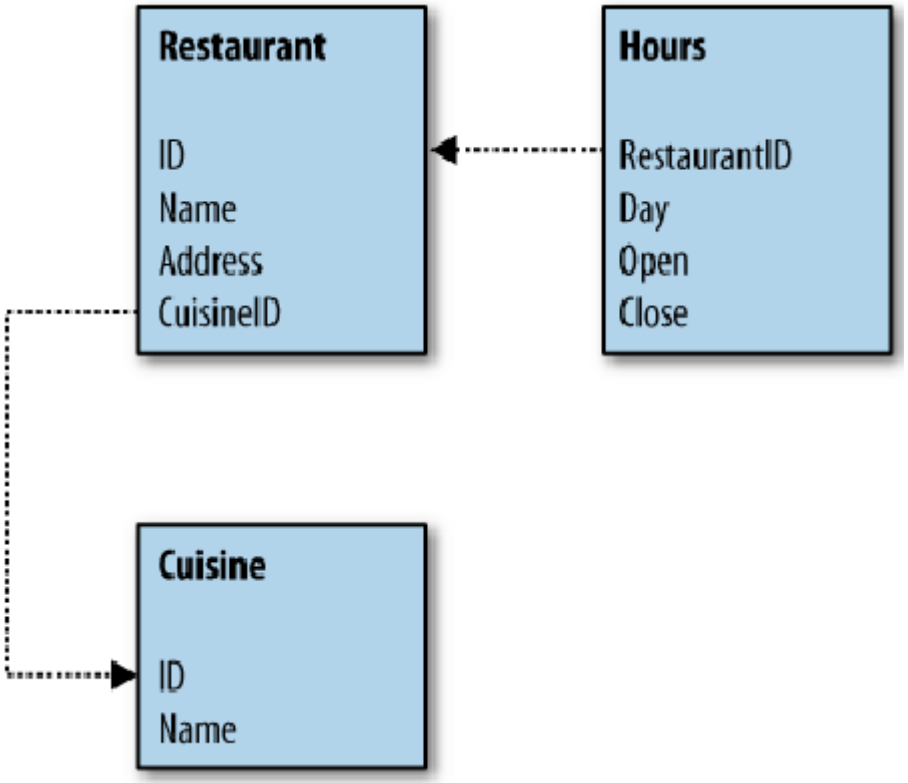
- O modelo tabular dá “impressão” de não possuir propriamente uma modelação
- Contudo, um olhar mais atento, revela que a localizção das linhas e das colunas fornece a cada dado um significado particular
- De facto, existe <u>semântica</u> numa simples tabela de dados
  - no exemplo anterior, o valor “Chinese” na coluna “Cusine”, revela, por exemplo que:
    - O restaurante Wok This Way serve comida chinesa.
- Este modelo é suficiente para consulta por utilizadores, mas demasiado limitado e rígido para ser manipulado por aplicações de software.

Métodos Tradicionais de Modelação de Dados

- Modelo Relacional
  - Baseado na lógica *First-Order Predicate*, proposta por Edgar Codd em 1969.
  - O modelo relacional, tem origem nas bases de dados relacionais (SQL Server, MySQL, Oracle, etc.) e permite muito maior flexibilidade na modelação dos dados.
  - Este modelo usa o modelo tabular como base, dispondo os dados em linhas (tuplos) numa tabela, mas permitindo o estabelecimento de relações entre diferentes tabelas.
  - Por conseguinte, é um modelo que oferece muito maior capacidade de manipulação dos dados a ferramentas de software.

Modelo Relacional (i)

- Exemplo de modelação



Modelo Relacional (ii)

- Exemplo de dados

Restaurant

ID	Name	Address	Price	CuisineID
1	Deli Llama	Peachtree Rd	$	1
2	Peking Inn	Lake St	$$$	2

Cuisine

ID	Name
1	Deli
2	Chinese
3	Thai
4	Fast Food

Hours

RestID	Day	Open	Close
1	Mon	11	16
1	Tue	11	16
1	Wed	11	16
1	Thu	11	19
1	Fri	11	20
2	Thu	5	22
2	Fri	5	23
2	Sat	5	23

Modelo Relacional (iii)

- No modelo do exemplo é patente uma maior explicitação da semântica.
- Os significados são descritos através do *schema*.
  - É possível verificar imediatamente que existem vários tipos de entidades:
    - Restaurante, Tipo de Cozinha, Horários
  - E que existem relações especificas entre elas.
- O significado dos dados é dado pelo significado intrínseco dos nomes das tabelas e das colunas.
- Apesar de a BD continuar a "<u>não saber</u>" o que é um restaurante, é possível perguntar por restaurantes que possuem um conjunto de características.

Modelo Relacional (iv)

- O modelo relacional é muito bom em cenários em que os dados são bem conhecidos e possuem uma estrutura estável
  - ou seja, conhece-se previamente o modelo de dados e as operações a executar sobre o mesmo
  - este conhecimento possui um tempo de vida longo
- No cenário global da web, estas assunções não são válidas
  - a estrutura dos dados está sempre a mudar assim como o conjunto de operações sobre os mesmos
  - aqui, os programadores nunca sabem com que tipo de dados vão contar e de que forma os utilizadores os vão querer usar

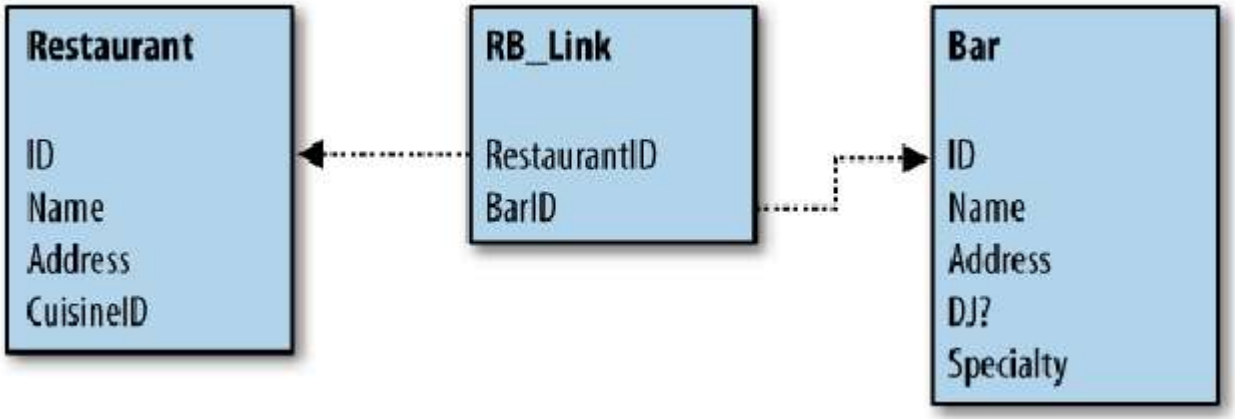
Modelo Relacional (v)

- Surgindo novos dados

Bar	Address	DJ	Specialty drink
The Bitter End	14th Ave	No	Beer
Peking Inn	Lake St	No	Scorpion Bowl
Hammer Time	Wildcat Dr	Yes	Hennessey
Marquis de Salade	Main St	Yes	Martini

Modelo Relacional (vi)

- Integração dos novos dados
  - Supondo que existem restaurantes que possuem bares e se pretende integrar os dados de forma coerente sem alterar a estrutura existente
  - Uma possibilidade é a ligação das 2 tabelas de restaurantes e bares através de uma tabela intermediária

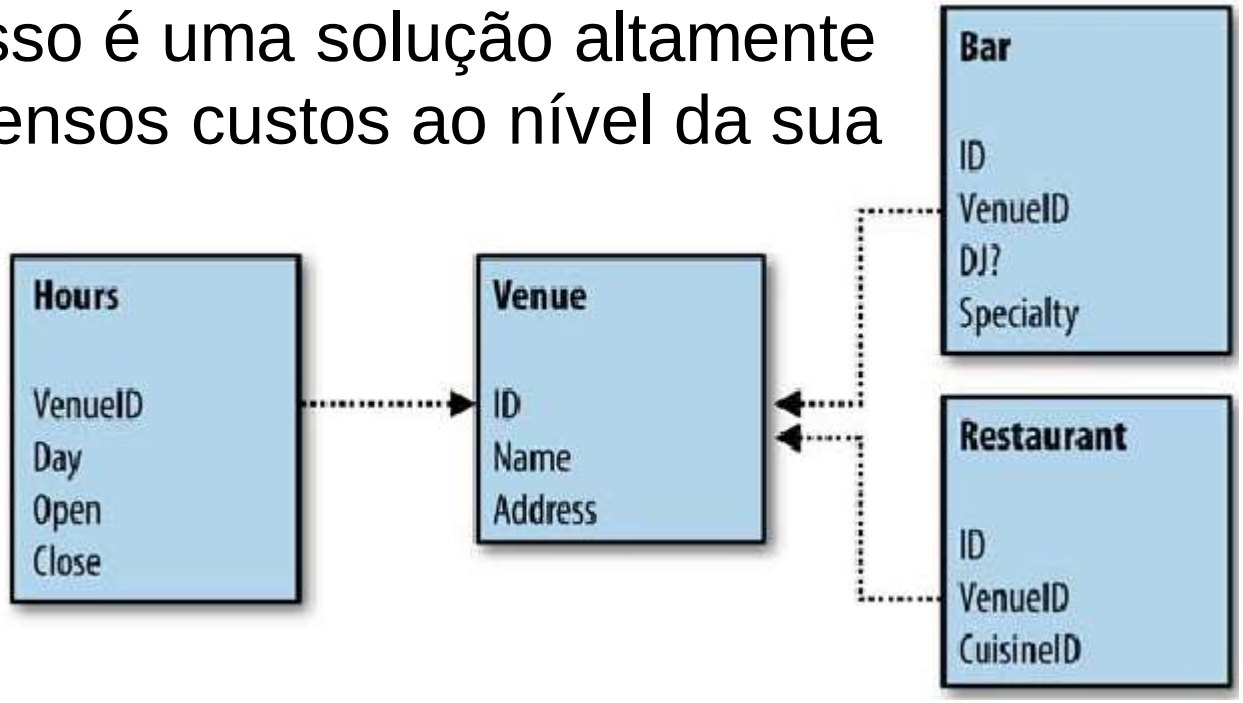


Modelo Relacional (vii)

- Integração de novos dados
  - A solução apresentada é pouco intrusiva
  - Mas também pouco eficiente
  - Exemplo:
    - Se se pretender saber os estabelecimentos existentes num determinado lugar, tem de se pesquisar as 2 tabelas de restaurantes e bares
    - A informação sobre o lugar pode encontrar-se redundante

Modelo Relacional (viii)

- Integração de novos dados
  - Uma solução mais eficiente, passa pela normalização dos dados
  - Isto implica a alteração do modelo de dados já existente e por isso é uma solução altamente intrusiva com imensos custos ao nível da sua reprogramação



Modelo Relacional (ix)

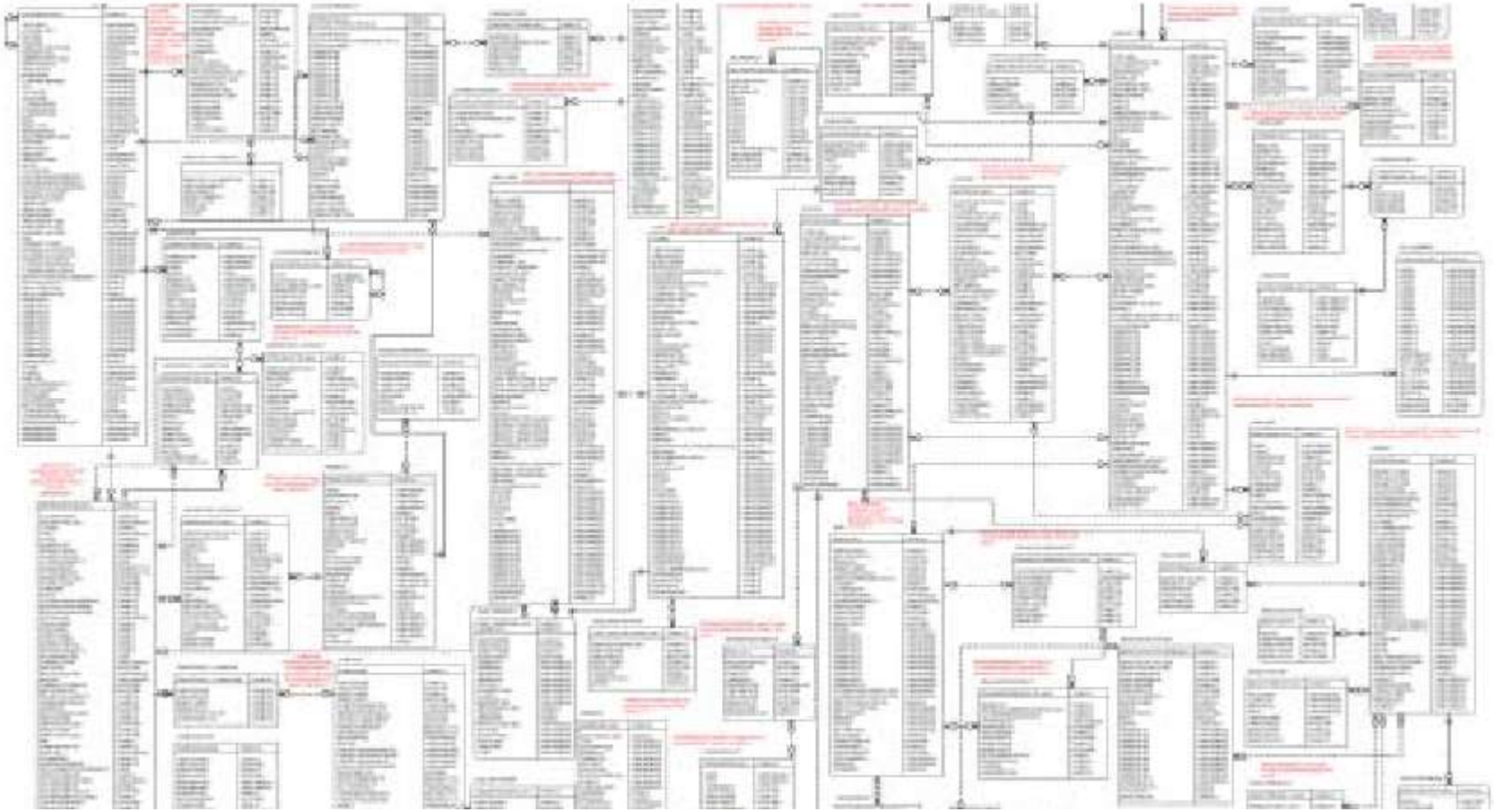
- Algumas conclusões sobre o uso do modelo relacional para integração de dados em cenários muito dinâmicos como a web:
  - A adoção de métodos pouco intrusivos levam rapidamente à ineficiência dos modelos de dados
    - Neste caso, os modelos vão crescendo até se tornar impossível a sua gestão, manutenção e utilização
  - A adoção de métodos intrusivos, mais eficientes, como refazer os modelos e a sua programação tornam-se impossíveis de realizar, pelo recursos necessários para tal

Modelo Relacional (x)

- Complexidade
  - Para além das questões levantadas antes, coloca-se ainda o problema da complexidade dos modelos relacionais que pode em muitos casos, só por si, inviabilizar qualquer das soluções propostas antes
  - Veja-se os exemplos de modelos relacionais de dados existentes em sistemas de CRM (*Customer Relationship Management*) e ERP (*Enterprise Resource Planning*)
    - Para estes sistemas são contratadas empresas de consultadoria dedicadas à gestão do modelo de dados

Modelo Relacional (xi)

- Complexidade – Exemplo

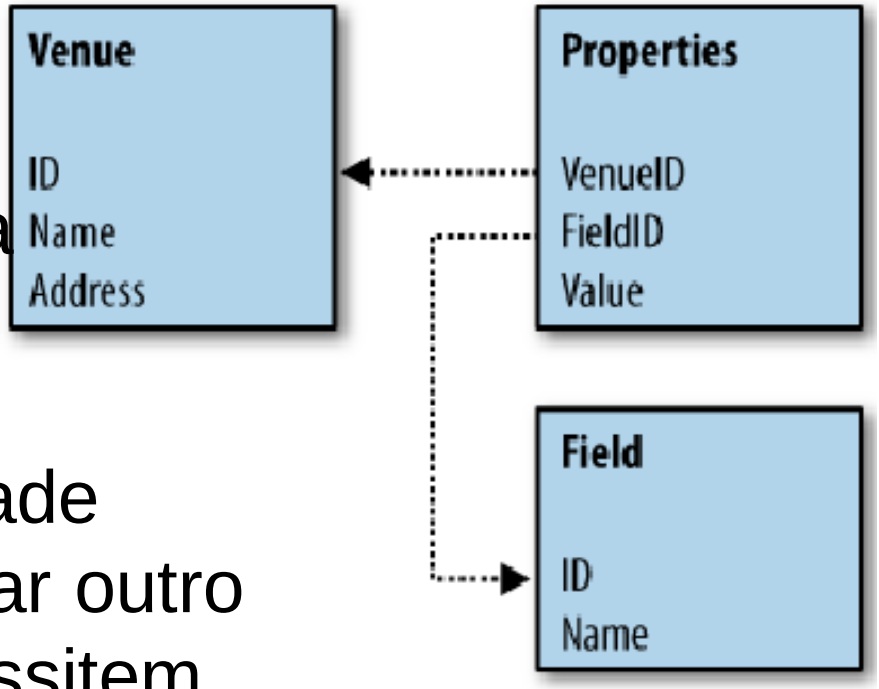


Modelo Relacional (xii)

- Hipotética solução
  - Seria possível conceber de início um modelo de dados suficientemente flexível para acomodar tipos de dados em permanente mudança?
  - E que, ao mesmo tempo, mantivesse um certo nível de legibilidade?
  - Proposta:
    - O modelo de dados poderia ser concebido da seguinte forma:
      - A Entidade central seria o “Lugar”
      - Esta <u>Entidade</u> seria caraterizada por um conjunto de <u>Propriedades</u>, com <u>Campos</u> parametrizáveis

Modelo Relacional (xiii)

- Proposta
  - [Lugar – Propriedades – Campo]
  - Em geral, esta proposta não é recomendável, pois perde muita da normalização conseguida e certamente degrada a performance
  - A vantagem é a flexibilidade necessária para acomodar outro tipo de lugares que necessitem de propriedades diferentes



Modelo Relacional (xiv)

- Proposta
  - Tabelas de Dados

Venue		
ID	Name	Address
1	Dell Llama	Peachtree Rd
2	Peking Inn	Lake St
3	Thai Tanic	Branch Dr

Properties		
VenueID	FieldID	Value
1	1	Dell
1	2	$
2	1	Chinese
2	2	$$$
2	3	Scorpion Bowl
2	4	No

Field	
ID	Name
1	Cuisine
2	Price
3	Specialty Cocktail
4	DJ?

Modelo Relacional (xv)

- Proposta
  - Novos Dados

Venue		
ID	Name	Address
1	Dell Llama	Peachtree Rd
2	Peking Inn	Lake St
3	Thai Tanic	Branch Dr

Properties		
VenueID	FieldID	Value
1	1	Dell
1	2	$
2	1	Chinese
2	2	$$$
2	3	Scorpion Bowl
2	4	No
3	5	Yes
3	6	Jazz

Field	
ID	Name
1	Cuisine
2	Price
3	Specialty Cocktail
4	DJ?
5	Live Music
6	Music Genre

Modelo Relacional (xvi)

- Proposta
  - Esta proposta consiste na verdade na implementação de um modelo que não é novo:
    - O modelo Chave–Valor (*Key–Value*)
  - No limite, tudo pode ser considerado como propriedade parametrizável da entidade “Lugar” e, neste caso, pode-se também juntar o nome e endereço

Properties		
VenueID	FieldID	Value
1	1	Deli
1	2	$
1	7	Deli Llama
1	8	Peachtree Rd
2	1	Chinese
2	2	$$$
2	3	Scorpion Bowl
2	4	No
2	7	Peking Inn
2	8	Lake St
3	5	Yes
3	6	Jazz
3	7	Thai Tanic
3	8	Branch Dr

Field	
ID	Name
1	Cuisine
2	Price
3	Specialty Cocktail
4	DJ?
5	Live Music
6	Music Genre
7	Name
8	Address

Modelo Relacional (xvii)

- Proposta
  - Na verdade, o modelação feita até ao momento ainda não é uma modelação chave-valor “pura”
  - Daí, que a relação entre as tabelas Campo e Propriedades só é conhecida através do conhecimento contido na lógica do query do tipo *Join*
  - O passo seguinte é então criar um modelo chave-valor puro e dessa forma explicitar na tabela todo o conhecimento

Semântica dos Dados

- Modelo Chave-Valor

VenueID	Field	Value
1	Cuisine	Deli
1	Price	$
1	Name	Deli Llama
1	Address	Peachtree Rd
2	Cuisine	Chinese
2	Price	$$$
2	Specialty Cocktail	Scorpion Bowl
2	DJ?	No
2	Name	Peking Inn
2	Address	Lake St
3	Live Music?	Yes
3	Music Genre	Jazz
3	Name	Thai Tanic
3	Address	Branch Dr

Semântica dos Dados

- Seguindo este modelo (chave-valor), cada dado é descrito diretamente através da propriedade que o define.
- Por este processo, as relações semânticas, que previamente eram inferidas dos nomes das tabelas e das colunas, são agora dados dentro da própria tabela.
- Esta é a essência da modelação semântica de dados: esquemas flexíveis, onde as relações são descritas pelos próprios dados.

Semântica dos Dados

- Metadados
  - Um dos desafios de usar dados relacionais de outrem é descobrir como é que as várias tabelas se relacionam entre si
  - Esta informação – dados acerca dos dados – é conhecida por metadados e consiste em conhecimento sobre como os dados podem ser usados
  - Este conhecimento é geralmente expresso explicitamente através da definição de chaves estrangeiras no modelo de dados, e/ou implicitamente através da lógica dos queries
  - Frequentemente, os dados são arquivados, publicados ou partilhados sem estes metadados
  - Quando se pretende um dia fazer a sua reutilização, a descoberta destas relações torna-se um exercício "*deveras interessante*"
  - Utilizando o modelo referido atrás, os metadados sobre o esquema relacional, no qual se descreve todas as colunas que descrevem uma entidade, passaram a fazer parte dos próprios dados