

# Conhecimento e Raciocínio

## Aula 5

### Raciocínio Baseado em Casos

---

*Viriato A.P. Marinho Marques*

DEIS - ISEC

2019 / 2020



# 6. CBR - Links Importantes

---

<http://www.ai-cbr.org/>

<http://www.ai-cbr.org/classroom/cbr-review.html> - Conclusions

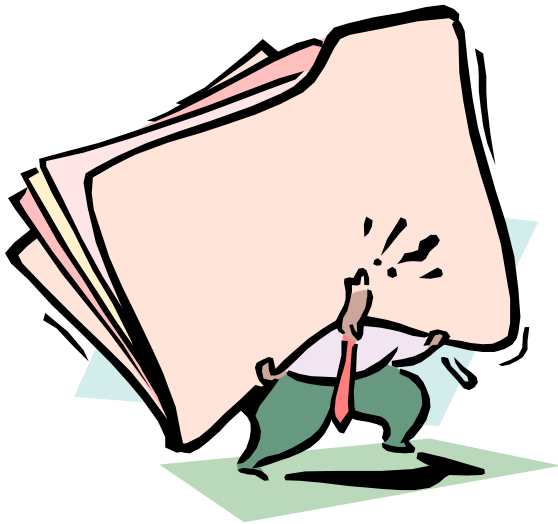
(resumo do livro de Ian Watson)

<http://www.cbr-web.org/>

(Universidade de Kaiserslautern)

<http://www.aic.nrl.navy.mil/~aha/research/case-based-reasoning.html>

(David Aha - USA Navy)



Há imensa informação na Web mas estes links, fundamentais, proporcionam ligação a (virtualmente) tudo o que é importante...

# 6.1 CBR - Introdução

---

## 6.1 Introdução

O Raciocínio Baseado em Casos (*Case Based Reasoning* ou CBR) é um paradigma (modelo) utilizado na implementação de muitos SPs:

- Baseia-se na representação de casos (ocorrências contextualizadas) que representam recordações (experiências passadas)
- Cada caso é composto por uma descrição e pela respectiva solução
- A resolução de um novo problema consiste em encontrar um caso passado semelhante ao actual e apresentar a sua solução directamente ou adaptada ao novo contexto

1º Sistema: CYRUS, Universidade de Yale, EUA, J.Kolodner, 1983

- Resposta sobre viagens do Secretário de Estado Cyrus Vince
- Baseado no trabalho R.Schank sobre organização dinâmica da memória

Europa: PATDEX, Universidade de Kaiserslautern, Alemanha, 1991

- Diagnóstico de falhas em sistemas complexos

# 6.1 CBR - Introdução

---



Vantagens dos SPs baseados no paradigma CBR:

- Funcionam de forma mais consentânea com o raciocínio pericial, i.e., baseiam-se em experiência passada
- A aprendizagem é simples: baseia-se na adição de "casos relevantes" à Base de Conhecimento (*Case Library*)
- A sua implementação evita a fase de extracção de conhecimento (pode ser essencialmente baseada em documentos que descrevem casos passados)

Desvantagens:

- Dificuldade na adaptação de casos: normalmente é realizada por regras (outras possibilidades consoante o domínio). Contudo, é difícil por natureza e assunto de investigação

Especialmente indicados para:

- Domínios que mudam com facilidade, dinâmicos
- Necessidade de aprendizagem constante e automática
- Domínios complexos e/ou mal compreendidos (caso típico: CLAVIER)
- Domínios que dependam claramente de experiências passadas (p.e. aconselhamento jurídico baseado em acções já julgadas)

# 6.1 CBR - Introdução

---

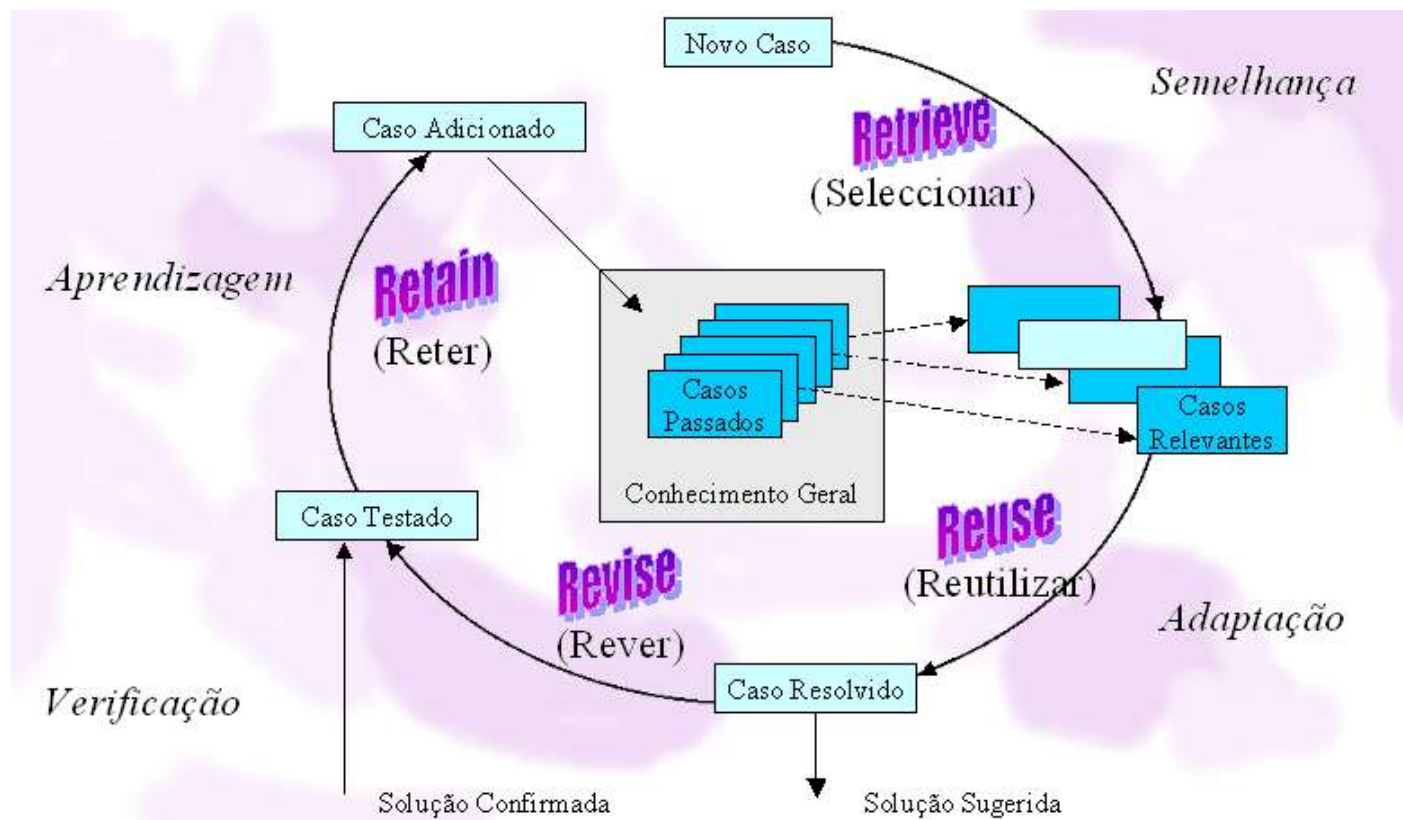
	Sistemas Baseados em Regras	Raciocínio Baseado em Casos
<b>Domínio</b>	Bem delimitado, bem compreendido, com alguma base, estável ao longo do tempo	Extenso, vago, mal compreendido, base teórica pobre, dinâmico ao longo do tempo
<b>Representação do Conhecimento</b>	Factos e regras de produção IF...THEN	Casos
<b>Sistema proporciona</b>	Respostas inferidas do zero	Respostas baseadas em acontecimentos precedentes
<b>Explicação por</b>	Traço das regras disparadas	Casos passados seleccionados
<b>O sistema pode aprender</b>	Geralmente não, regras modificadas manualmente, quando muito indução de regras a partir de exemplos (CBR ou regras ?)	Sim, por simples aquisição de novos casos (o sistema aprende com o seu uso)



# 6.2 Ciclo de Aamodt & Plaza

## 6.2 Ciclo de Aamodt & Plaza

A operação de um sistema CBR é tipicamente descrita pelo ciclo de Aamodt e Plaza



# 6.2 Ciclo de Aamodt & Plaza

---

Ciclo de 4 fases:

- **Retrieve:** Procurar o caso ou os casos mais semelhantes ao actual
- **Reuse:** Reutilizar a informação que esse caso contém para resolver o problema actual. A solução pode ser exactamente a mesma ou ter de ser adaptada.
- **Revise:** Testar se a solução proposta é válida. Se inválida, proceder à sua correcção
- **Retain:** Reter o conhecimento contido no caso actual se contiver algo de diferente; se ensinar uma lição

Terminologia:

- Biblioteca de Casos (*Case Library*): A estrutura do sistema onde estão armazenados os casos (vulgarmente uma Base de Dados Relacional, em especial nos *shells* comerciais)
- *Seed Cases*: Casos iniciais carregados no sistema para início de operação



## 6.2.1 Retrieve

---

### 6.2.1 Retrieve

1. Seleccionar um conjunto de casos (possivelmente) relevantes para a solução do problema actual.
2. Seleccionar o melhor ou melhores de entre eles.

#### Fase 1

Vulgarmente os casos contêm pares atributo-valor, p.e. (Tensão,220V)...

Nesta fase são seleccionados casos com contornos (perfil) que parece semelhante ao do caso actual por terem índices iguais.

Contudo, a indexação para CBR difere da indexação vulgar de RDBs:

- RDBs: Os índices são de natureza funcional (surtem em consequência do desenho e existem frequentemente em chaves primárias e candidatas)
- CBR: Os índices visam a rápida identificação dos casos cujos atributos, contornos, sugerem que sejam relevantes para a resolução do problema actual

*"Estamos preocupados em seccionar a biblioteca de casos em fragmentos conceptualmente úteis"*  
(Kolodner & Leake, 1996).





## 6.2.1 Retrieve

---

Para permitirem prever a utilidade dos casos na biblioteca, os índices:

- Têm de ser abstractos para funcionarem numa variedade de situações futuras
- Têm de ser suficientemente concretos para impedirem a selecção de casos que não serão úteis na resolução de um problema

### Fase 2

Habitualmente implementada pelo “algoritmo do vizinho mais próximo” (***nearest neighbor***) muito característico do paradigma CBR.

Calcula uma **semelhança global** entre o caso actual e cada um dos casos relevantes seleccionados na fase 1

Baseia-se nos valores dos atributos dos casos em consideração

Na sua versão mais simples:

- É calculada uma distância entre valores de atributos correspondentes;
- Estas distâncias são adicionadas;
- A sua soma é transformada numa medida de semelhança



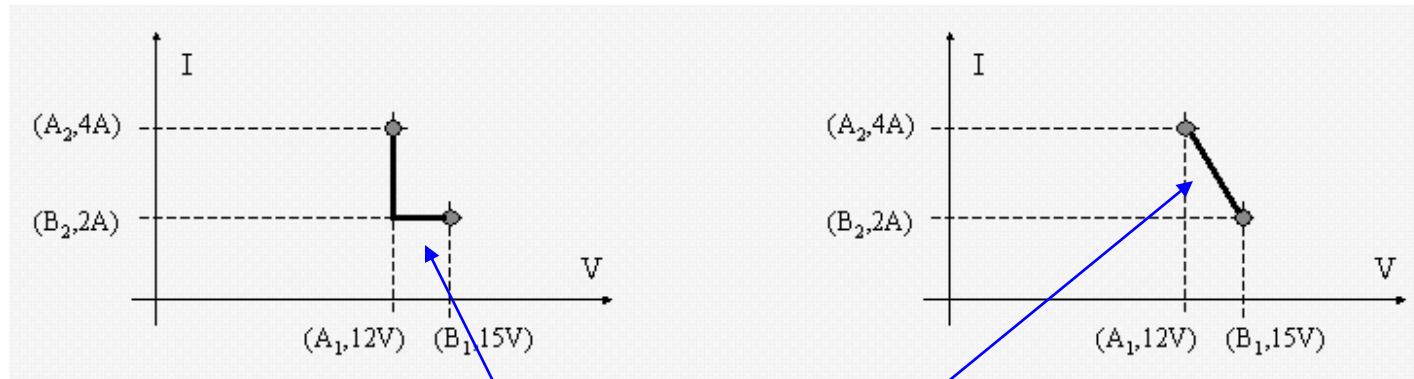
## 6.2.1 Retrieve

### Exemplo:

Considerem-se dois casos que envolvem dois transformadores

Atributos: Tensão de saída e Intensidade de corrente do secundário

Valores: **Caso 1:**  $A_1=12V$ ,  $A_2=4A$     **Caso 2:**  $B_1=15V$  e  $B_2=2A$



- A distância pode ser Linear ou Euclidiana (há muitas formas de cálculo de distâncias consoante os SPs, domínios, protótipos, etc.)
- Para calcular semelhanças entre casos as medidas têm de ser normalizadas:

## 6.2.1 Retrieve

---

Normalizando em relação aos valores máximos 15V e 4A obtém-se

$$A_{1N}=12/15=0.8 \quad B_{1N}=4/4=1$$

$$A_{2N}=15/15=1 \quad B_{2N}=2/4=0.5$$

As distâncias normalizadas lineares e euclidianas, DL e DE , serão então:

$$D_L = \frac{1}{2}(|A_{1N} - B_{1N}| + |A_{2N} - B_{2N}|) = \frac{1}{2}(0.2 + 0.5) = 0.35 \quad \text{Linear}$$

$$D_E = \sqrt{\frac{(A_{1N} - B_{1N})^2 + (A_{2N} - B_{2N})^2}{2}} = \sqrt{\frac{0.04 + 0.25}{2}} = 0.38 \quad \text{Euclidiana}$$

- Em CBR cada atributo é normalmente afectado por **factores de relevância**  $w_i$  (*weighting factors*)
- Expressam quanto esse atributo importante para a semelhança global

## 6.2.1 Retrieve

---

Como consequência:

- Na fórmula das distâncias os denominadores de valor 2 devem ser substituídos pelo somatório dos factores de relevância, de modo a que a distância global continue a situar-se no intervalo  $[0,1]$

### Distância

#### Linear para $n$ atributos

$$D_L(q, s) = \frac{\sum_{i=1}^n d(q_i, s_i) \times w_i}{\sum_{i=1}^n w_i}$$

$D_L$  Distância linear global entre casos

$D_E$  Distância euclidiana global entre casos

$q$  Caso actual

$s$  Caso da biblioteca

$q_i$  Valor do atributo de índice  $i$  no caso actual

$s_i$  Valor do mesmo atributo no caso da biblioteca

$w_i$  Factor de relevância para o atributo de índice  $i$

$d$  Função distância entre atributos correspondentes

### Distância

#### Euclidiana para $n$ atributos

$$D_E(q, s) = \sqrt{\frac{\sum_{i=1}^n [d(q_i, s_i)]^2 \times w_i}{\sum_{i=1}^n w_i}}$$

## 6.2.1 Retrieve

---

O valor da semelhança global,  $S$ , é normalmente definido como a diferença de uma **distância normalizada**,  $D$ , para 1 :

$$S = 1 - D$$

Variantes da expressão anterior (Richter et al., 1991):

- Se  $D$  for um valor não normalizado com um máximo conhecido  $\max(D)$

$$S = 1 - (D / \max(D))$$

- Se  $D$  puder assumir um valor infinitamente grande.

$$S = 1 - (D / 1 + D)$$

Para qualquer delas:

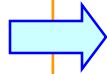
$$\begin{cases} S=1 & \text{se } D=0 \\ S=0 & \text{se } D=\max(D) \text{ ou } D=\infty \end{cases}$$



## 6.2.1 Retrieve

Atendendo à linearidade de  $D_L(q, s) = \frac{\sum_{i=1}^n d(q_i, s_i) \times w_i}{\sum_{i=1}^n w_i}$  e a  $S=1-D$ , temos que:

**Semelhança linear  
para  $n$  atributos  
(Forma Vulgar)**



$$S(q, s) = \sum_{i=1}^n s(q_i, s_i) \times w_i / \sum_{i=1}^n w_i$$

com  $s(q_i, s_i) = 1 - d(q_i, s_i)$

**Exemplo:**

Atr. / Valores:

**Caso 1:**  $A_1=12V$ ,  $A_2=4A$

**Caso 2:**  $B_1=15V$  e  $B_2=2A$

Normalizados:

**Caso 1:**  $A_1=0.8$ ,  $A_2=1$

**Caso 2:**  $B_1=1$  e  $B_2=0.5$

Relevâncias:

$w(A) = 2$ ,  $w(B) = 3$

$$D_L(q, s) = \frac{|0.8 - 1| \times 2 + |1 - 0.5| \times 3}{2 + 3} = \frac{1.9}{5} = 0.38$$

$$S = 1 - D = 1 - 0.38 = 0.62$$

$$S(q, s) = \frac{(1 - |0.8 - 1|) \times 2 + (1 - |1 - 0.5|) \times 3}{2 + 3} = \frac{1.6 + 1.5}{5} = 0.62$$

## 6.2.2 Reuse

---

### 6.2.2 Reuse

1. Propor a solução de um caso passado como solução para o caso actual (*null adaptation*)

OU

2. Adaptar a solução de um caso passado ao caso actual:
  - **Adaptação Transformacional:** são definidos operadores transformacional  $\{T\}$  - p.e. regras, algoritmos, expressões matemáticas - que guiam a adaptação da solução passada ao caso actual
  - **Adaptação Derivacional:** um método usado no passado para adaptar uma solução é reutilizado; a cada caso tem de ser associada alguma forma de descrição desses métodos

É utilizada em muitos SPs baseados em casos, dos quais alguns exemplos típicos são o CHEF (culinária), o JUDGE (jurídico) e o CLAVIER (cura de compósitos em autoclave):

## 6.2.2 Reuse

---



Exemplos de **adaptação**:

- **Por Reinstanciação**

- CLAVIER - Na substituição de peças de certo formato por outras que provavelmente permitirão uma cura bem sucedida
- CHEF - Na substituição de alguns elementos de um prato por outros, p.e. *ervilhas* por *legumes* ou *carne de vaca* por *frango*

- **Guiada por Parâmetros**

- JUDGE - No agravamento de uma sentença por roubo caso tenha sido praticado à mão armada

- **Guiada por um Modelo do Domínio**

- CASEY (Koton, 1989) - diagnóstico de doenças cardíacas (recorre a um modelo do domínio chamado *Heart Failure Program*)

A adaptação é assunto de investigação e discussão.

As soluções são muito variadas.

CLAVIER - Os autores constataram que os utilizadores preferiam fazer as adaptações manualmente, usando o sistema apenas para as validar: "*Uma das nossas conclusões foi que a adaptação automática de casos não era possível*".

## 6.2.3 Revise

---

### 6.2.3 Revise

#### 1. Avaliação da solução proposta

- Por um **professor externo**:

Manual

- Diagnóstico: um técnico ou médico confirma ou não a solução proposta pelo sistema

Automática

- Planeamento: CLAVIER - a equipa confirma se a carga proposta foi bem sucedida ou não

- Por um **processo de simulação**:

- CHEF - As receitas culinárias propostas pelo SP são aplicadas a um simulador que tenta validá-las

#### 2. Eventual **reparação** dessa solução

- Diagnóstico: o diagnóstico proposto é modificado pelo utilizador que pode também comunicar um novo ou escolher um já existente mas não associado ao contexto actual
- CLAVIER: se a carga proposta for indicada como "mal sucedida" pelo utilizador (aqui é impossível reparar "completamente" a solução porque não se sabe "como poderia ser bem sucedida") 17

## 6.2.4 Retain

---

### 6.2.4 *Retain*

1. Que conhecimento reter ?
2. Quando o reter ?
3. Como o reter?

Há que ter em conta:

- O processo de representação dos casos
- O modo como a aprendizagem é realizada

#### 6.2.4.1 Representação de Casos - Protótipos / Modelos Académicos

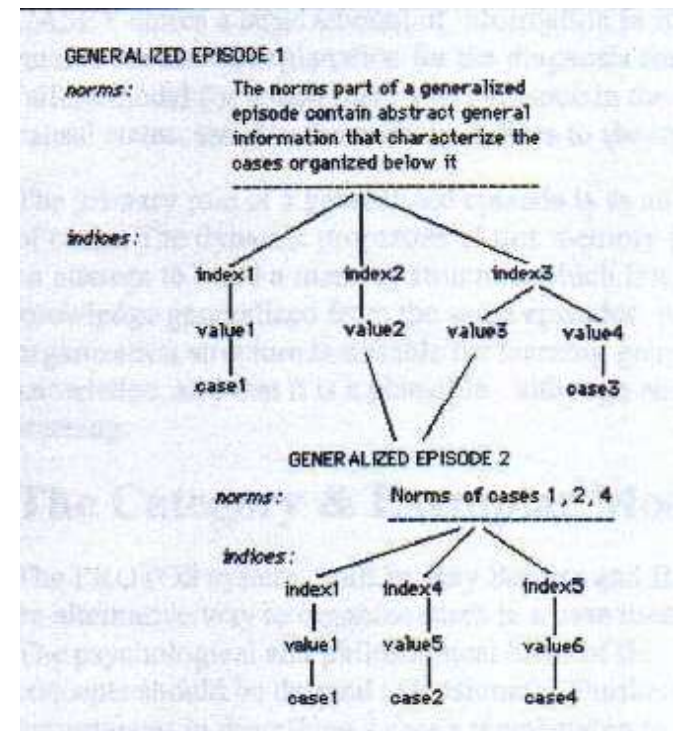
- Trabalhos pioneiros de (Schank,R., 1982) e (Kolodner,J., 1983):
  - Modelo de **Memória Dinâmica** (CYRUS, CASEY)
- (Bareiss et al., 1988):
  - Modelo **Categoria-Exemplo** (PROTOS)



## 6.2.4 Retain

- **Modelo de Memória Dinâmica (CYRUS, CASEY)**

- Organizar os casos que partilham propriedades semelhantes, numa estrutura mais geral chamada “episódio generalizado”
- Um episódio generalizado contém 3 tipos de objectos: Normas, casos e índices:
  - Normas são as propriedades comuns a todos os casos do mesmo episódio generalizado;
  - Índices apontam um episódio generalizado mais específico, ou um caso;
- Episódios generalizados e casos compõem uma rede que, quer para a pesquisa de casos semelhantes, quer para a actualização da biblioteca, é percorrida com vista à localização do caso mais próximo
- A inserção de um novo caso pode levar à criação de um novo episódio generalizado, se esse caso tiver propriedades comuns com o seu vizinho mais próximo

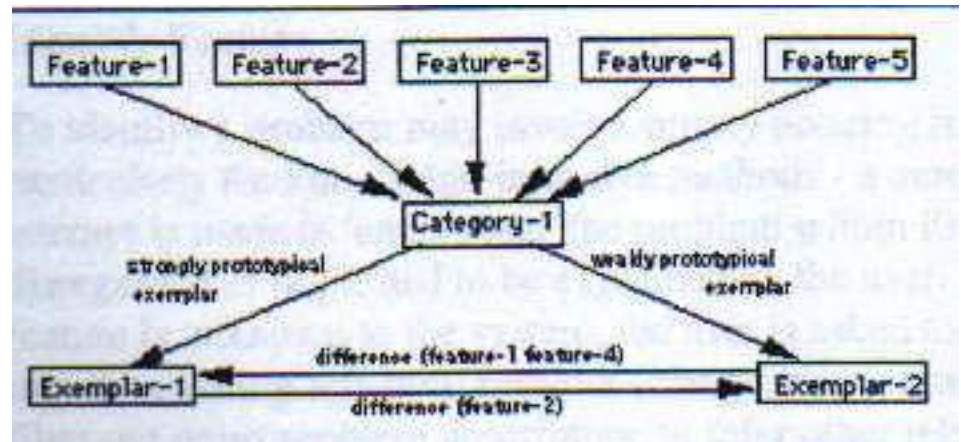


## 6.2.4 Retain

- **Modelo Categoria-Exemplo (PROTOS)**

- Os casos são designados por “exemplos”
- A representação é efectuada à custa de propriedades, categorias, exemplos (casos), e índices
- Os índices são de 3 tipos:
  - Ligando um conjunto de propriedades (“recordações”) a uma categoria ou a um exemplo;
  - Ligando uma categoria aos seus exemplos;
  - Ligando dois exemplos entre eles (este tipo de ligação só existe entre exemplos vizinhos que diferem entre eles por um pequeno número de recordações e que estão todos associados à mesma categoria)

- Os exemplos de cada categoria são ordenados por grau de tipicidade e as categorias encontram-se interligadas numa rede semântica



## 6.2.4 Retain

---

### 6.2.4.2 Representação de Casos - *Shells* Comerciais

- Não utilizam qualquer destes tipos de representação
- Utilizam RDBs
- Algumas têm ferramentas de indução (ID3 / Data-Mining) que permitem a criação de árvores de decisão a partir de exemplos:
  - **Eager-Learning** (Casos como base para ID3): Aprendizagem "antecipada", i.e., o conceito generalizado cria-se no momento em que os exemplos de treino são apresentados (redes neuronais, ID3)
  - **Lazy-Learning ou Instance Based Learning** (CBR na essência): Os exemplos são arquivados para uso mais tarde (daí o "lazy"), por selecção dos vizinhos mais próximos (*k-nearest-neighbor*) (David Aha, USA Navy)
- O caso actual deverá ser inserido como **novo caso** ?
- Deverão ser feitas apenas algumas "**actualizações**" na base de dados ?
  - A decisão é tomada em função do conteúdo do caso actual e do conteúdo dos casos que já figuram na biblioteca;
  - A aprendizagem realiza-se quer por erro, quer por sucesso, pelo que também é importante considerar o desfecho obtido no processamento



## 6.2.4 Retain

---

Devido à variedade e complexidade dos domínios não é possível definir regras estritas.

Segundo Schank, 1996:

“Todas as pessoas se guiam pelo mesmo ‘script’ quando almoçam num McDonalds porque todos os McDonalds são basicamente iguais. Contudo, se alguém jantar no Taillevent pela primeira vez, começará por se guiar pelo ‘script’ respeitante ao melhor restaurante da sua terra. E em seguida irá descobrir que alguns desses princípios também funcionam bem ali (por exemplo, ser-lhe-á dada uma lista de vinhos e esperado que escolha um vinho que deverá provar antes de ser servido). Porém, quando alguma destas regras falhar (por exemplo, quando devolver o chocolate que veio com o café dizendo que não o pediu) ... então deverá começar a aquisição de um novo caso.”

Segundo Kolodner & Leake, 1996:

"Se o que é diferente numa nova situação ensina qualquer coisa que não poderia facilmente ser inferida a partir dos casos já gravados, então é útil gravar essa nova situação como novo caso"



## 6.2.4 Retain

---

### Princípios orientadores:

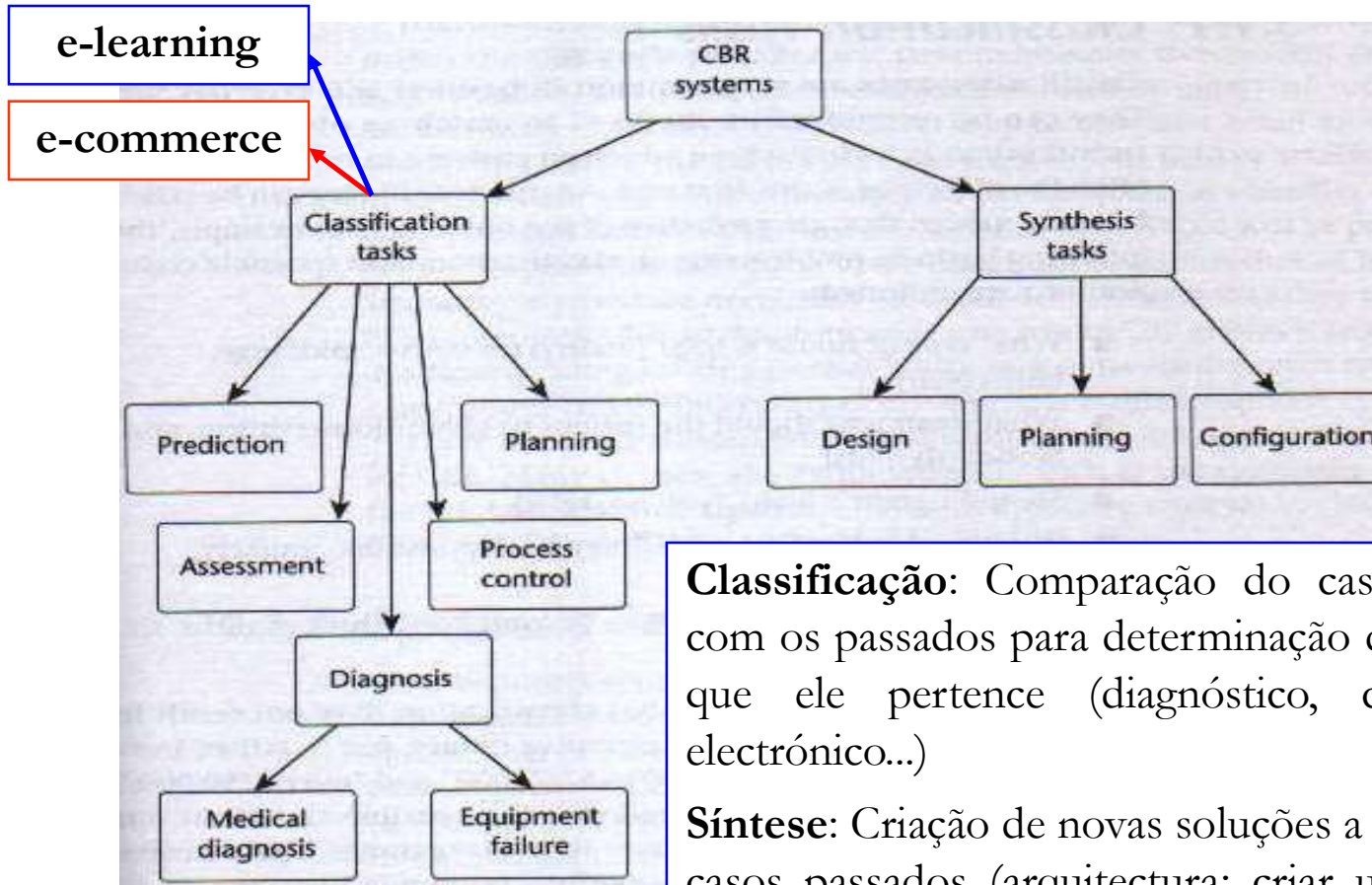
- Num sistema CBR a aprendizagem pode ser realizada
  - Por **inserção** de casos novos
  - Por **actualização** (tipificação ou generalização) de casos passados
  - Por actualização de **factores de relevância** (os coeficientes  $w$  das fórmulas do *nearest-neighbor*)
  - Por **indução** de árvores de decisão geradas por ID3 (ou algoritmo semelhante) quando na biblioteca de casos forem realizadas modificações
- Um caso deve ser inserido
  - Se o seu perfil (índices) for diferente
  - Se a solução gerada no Reuse não tiver sido directamente utilizada:
    - Por ter sofrido modificações manuais ou automáticas na fase de Revise
    - Por ter falhado tendo sido indicada uma solução "completamente nova"
  - Um caso pode ser inserido como exemplo "positivo" (i.e. "a este tipo de caso corresponde a solução X") ou "negativo" (i.e. "para este tipo de caso a solução Y não é válida". **Exemplo:** CLAVIER que regista cargas que falharam: estes exemplos são usados para invalidar soluções criadas manualmente)





# 6.3 CBR - Classificação

## 6.3 Classificação de Sistemas CBR (Althoff et al. 1995)



**Classificação:** Comparação do caso actual com os passados para determinação do tipo a que ele pertence (diagnóstico, comércio electrónico...)

**Síntese:** Criação de novas soluções a partir de casos passados (arquitectura: criar uma casa com base em divisões de outros projectos)

# 6.3 CBR - Classificação

---

## Classificação

- **Diagnóstico:** Diagnóstico médico e falhas em equipamento (CASEY, PROTOS, PATDEX, Ariane Espace, ...)
- **Predição:** Previsão de falhas de equipamento (Manutenção Predictiva) e previsão meteorológica ([CBRMeteorologia.pdf](#))
- **Cálculo de Riscos:** Empréstimos bancários, seguradoras, custo previsível de projectos
- **Controlo de Processos:** Controlo de processos de fabrico (**Exemplo: SHAI Esteem - Set points na Indústria Siderúrgica**)
- **Planeamento:** Processos de fabrico (**CLAVIER**), planeamento de viagens
- **Comércio Electrónico:** Pesquisa de produtos semelhantes ao perfil descrito pelo cliente (os contornos podem ser ponderados) (**Exemplo: Kaidara-commerce**)
- **Ensino à Distância:** Pesquisa de exemplos adequados à matéria leccionada e ao perfil do aluno (em investigação)

## Síntese

- Criação de um novo produto (**arquitectura: ARCHIE**), plano ou configuração a partir de partes de outros anteriores bem sucedidos

# 6.4 Biblioteca e Casos Iniciais

---

## 6.4 Biblioteca e Casos Iniciais (*Seed Cases*)

Para que um SP-CBR inicie a sua operação tem de dispor de casos iniciais (*seed cases*). Segundo Watson:

- Os casos devem ser representativos e bem distribuídos:
  - **Representativos:** descritos por um conjunto de atributos importantes e de valores típicos;
  - **Bem distribuídos:** o conjunto de casos deve cobrir uma boa percentagem das ocorrências que se prevêem

Um sistema só é bem aceite pelos seus utilizadores quando

- Consegue resolver, no início de operação, cerca de 80% dos problemas que lhe são postos (Exemplo: se os casos de que se dispõe puderem ser classificados em 6 tipos e se se verificar que 3 destes tipos ocorrem 80% das vezes, então pelo menos estes deverão ser carregados inicialmente)

O número de casos iniciais e o número de casos ao longo do tempo de vida varia muito

# 6.4 Biblioteca e Casos Iniciais

---

## Exemplos:

- CASEY: Para diagnóstico de doenças do coração, iniciou a operação com 24 casos
- CASELine: Usado na British Airways para diagnóstico de falhas em motores de Boeing 747. Em 1994 continha 200 casos
- CLAVIER: Esperava-se que o número de casos estabilizasse ao fim de dois anos em cerca de 300. Contudo não parou de crescer e continha, segundo (Mark et al., 1996), nessa data, cerca de 600, devido essencialmente a modificações do processo de fabrico e refinamento da Base de Conhecimento.
- CASSIOPE: Diagnóstico de falhas em motores da Boeing. Operava com 16.000 casos em 1995
- SQUAD: Para controlo de qualidade de software, iniciou com 25.000 casos
- ALFA: Sistema de previsão de cargas para centrais de produção de energia, operava em 1988 com cerca de 87.000 casos
- FL-CBR: Baseado em CBR e Lógica Difusa. Previsão meteorológica em aeroportos: Utiliza cerca de 300.000 casos fornecidos pelo Halifax International Airport, Nova Scotia, de 1961 a 1996, hora a hora





# 6.5 Alguns Protótipos Académicos

---

## 6.5 Alguns protótipos

### CHEF (Hammond, 1986)

Adaptação

- Receitas culinárias novas a partir de outras conhecidas
- Ilustra adaptação de casos (por reinstanciação):
  - Por exemplo, a partir de uma receita que contem “galinha” e “ervilhas”, o CHEF é capaz de criar uma nova receita baseada em “carne de vaca” e “bróculos”, ou seja, criando novas instâncias das classes “carne” e “legume”
  - Os resultados são submetidos a um módulo que os critica podendo obrigar a correcções (automáticas)
- Cada nova receita válida passa a constituir um novo caso

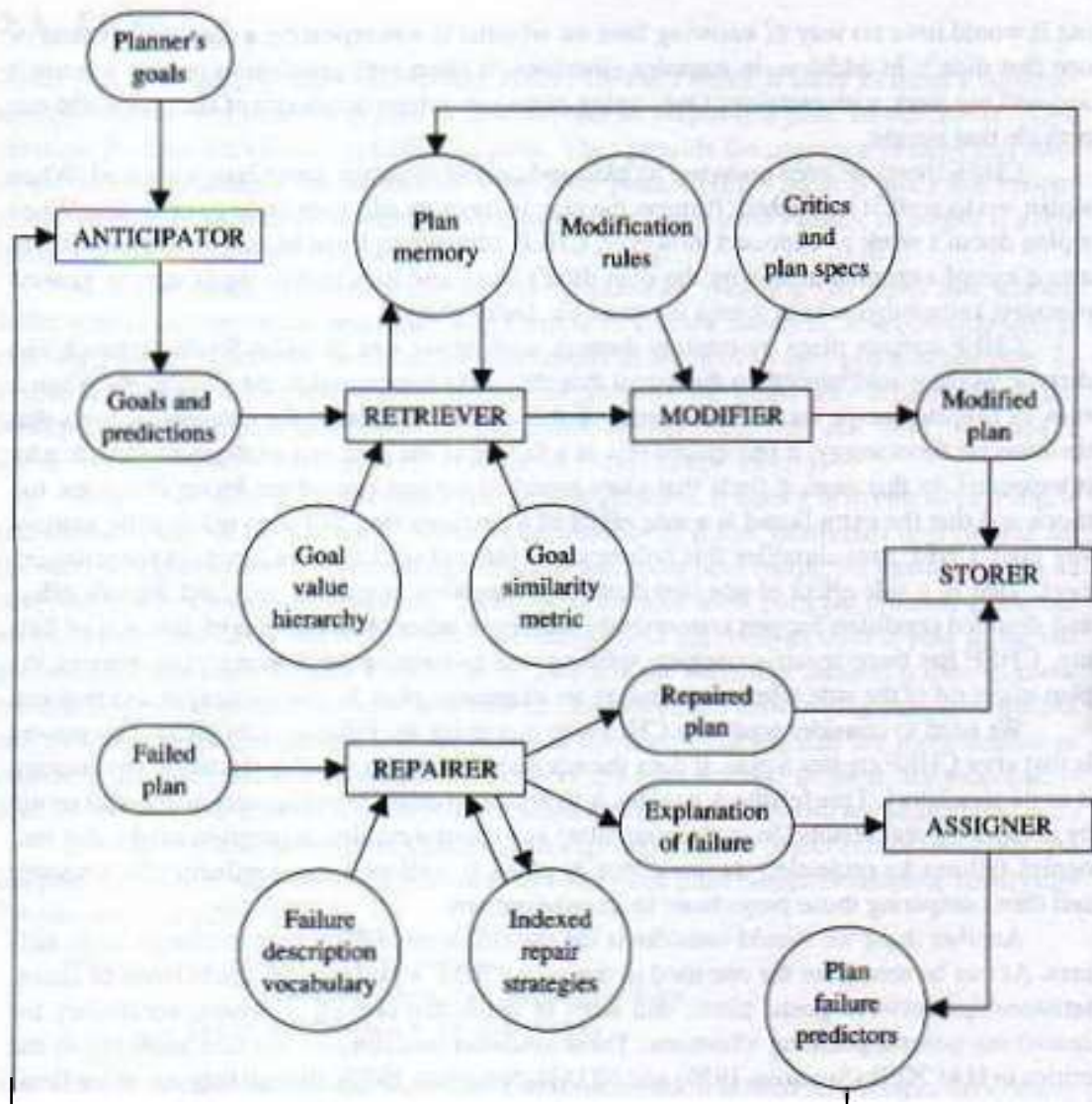
### ARCHIE (Pearce et al., 1992)

Design

- Arquitectura: acede a projectos anteriores (criados por outros arquitectos)
- Cada caso (projecto) contém informação acerca dos objectivos atingidos e restrições satisfeitas
- Para o novo projecto aponta factores a tomar em consideração
- Desenvolvido com base no Remind (Cognitive Systems, cessou em 1996)<sub>28</sub>



## 6.5 Alguns Protótipos Acadêmicos



Arquitetura  
do  
CHEF

## 6.5 Alguns Protótipos Académicos

---

### PROTOS (Porter & Bareiss, 1986)

Diagnóstico Médico

- Diagnóstico de doenças auditivas
- Classifica as doenças em função de sintomas, história clínica e análises
  - Foi treinado com 200 casos divididos por 24 categorias, obtendo uma taxa de acerto de 100% (Kolodner, 1996)
- Em memória os casos são organizados num modelo semelhante ao categoria-exemplo havendo vários tipos de *links* entre casos (p.e. ***difference-links*** que permitem distinguir diferenças subtis entre diferentes doenças)
- Inicialmente desenvolvido em Prolog

### CASEY (Koton, 1989)

Diagnóstico Médico

- Diagnóstico de doenças de coração
- 24 *seed cases* criteriosamente seleccionados:
  - Compostos por associações sintomas-diagnóstico
  - Detalhadamente descritos por uma longa série de atributos
- Dado um novo caso:
  - O CASEY começa por procurar o mais semelhante utilizando **factores de relevância dependentes do sintoma e do diagnóstico em consideração**
  - O caso seleccionado fornece uma primeira aproximação do diagnóstico

## 6.5 Alguns Protótipos Académicos

- Como os casos são descritos por uma extensa série de atributos, alguns deles podem conter diferenças significativas.
- Se tal acontecer, essas diferenças dão entrada num modelo detalhado chamado *Heart Failure Program* (Long, 1989)
- Este modelo realiza inferências que passam pela redução dos casos a estados patológicos básicos, de tal modo que casos aparentemente diferentes, podem afinal vir a revelar-se próximos (Kolodner, 1996)  
(**Exemplo:** uma ligação entre “ritmo cardíaco irregular” e “arteriosclerose”, pode vir a ser substituída por outra entre “tensão alta” e “arteriosclerose”)

Feature Name	Value for Old Case (David)	Value in New Case (Newman)
Age	72	65
Pulse rate	96	90
Temperature	98.7	98.4
Orthostatic change	absent	unknown
Angina	unstable	within-hours & unstable
Mean arterial pressure	107	99.3
Syncope	none	on exertion
Auscultation	murmur of AS	unknown
Pulse	normal	slow-rise
EKG	normal sinus & LV strain	normal sinus & LVH
Calcification	none	mitral & aortic

Extracto da  
descrição de 2  
casos no CASEY

# 6.5 Alguns Protótipos Acadêmicos

## PATDEX (Richter, 1991)

### Diagnóstico de Falhas

- Diagnóstico de falhas em equipamento complexo
- Existem duas versões (PATDEX-1 e PATDEX-2)
- PATDEX-1:
  - Um grafo liga casos entre si, de tal forma que o caso seguinte contém apenas mais um sintoma que o anterior;
- PATDEX-2:
  - Associa cada observação a cada diagnóstico através de uma **matriz de relevâncias** (coeficientes  $w_{ij}$  = relevância de cada observação  $i$  para cada diagnóstico  $j$ ).
  - O ajuste destes coeficientes é feito automaticamente
  - Estes  $w_{ij}$  figuram nas **equações de semelhança global** multiplicando semelhanças locais entre atributos

$$R = \begin{matrix} & \phi_1 & \phi_2 & \dots & \phi_m \\ \begin{matrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{matrix} & \begin{matrix} w_{11} \\ w_{21} \\ \vdots \\ w_{n1} \end{matrix} & \begin{matrix} w_{12} \\ w_{22} \\ \vdots \\ w_{n2} \end{matrix} & \dots & \begin{matrix} w_{1m} \\ w_{2m} \\ \vdots \\ w_{nm} \end{matrix} \end{matrix}$$

$$sim(Sit_1, Sit_2) = \frac{\alpha E}{(\alpha E + \beta C + \eta U_1 + \gamma U_2)}$$

$$E := \sum_{S_i \in E} w_{ij} \omega_i(a_{ik}, a_{il})$$

$$C := \sum_{S_i \in C} w_{ij} (1 - \omega_i(a_{ik}, a_{il}))$$

$$U_1 := \sum_{S_i \in U_1} v_{ij} (1 - \omega_i(a_{ik}, a_{il} J)) = card(U_1)$$

$$U_2 := \sum_{S_i \in U_2} w_{ij} (1 - \omega_i(a_{ik}, a_{il})) = \sum_{S_i \in U_2} w_{ij}$$



# 6.5 Alguns Protótipos Académicos

---



## JUDGE (Bain, 1986)

Domínio Jurídico

- Sentenças criminais (assassinio, assaltos, etc.)

## HYPO (Ashley, 1988)

- Direitos de autor e patentes: gera argumentos para defesa ou acusação baseado em casos anteriores

## PERSUADER (Sycara, 1987)

- Resolução de conflitos laborais
- Propõe novos contratos por adaptação de contratos usados por companhias semelhantes
- Pode gerar contratos novos se não encontrar nenhum relevante, através de um sistema de planeamento

## JULIA (Hinrichs, 1989)

Design

- Projecto de ementas
- Usa métodos de adaptação para criar novas ementas que respeitem as restrições actuais



## 6.5 Alguns Protótipos Académicos

---

### Projecto INRECA e CASUEL - 1994

- O CASUEL não é um SP mas uma "linguagem de representação de casos" - *Common Case Representation Language*: tentativa de estabelecimento de um standard para troca de informação entre sistemas de classificação e diagnóstico baseados em casos
- Nasceu do projecto INRECA - *Induction and Reasoning form Cases* - financiado pelo ESPRIT III (CEE) de signatários Universidade de Kaiserslautern, tec:Inno (hoje Empolis), AcknoSoft (hoje Kaidara) e IMS - *Irish Multimedia Systems*
- Do INRECA resultaram dois CBR shells:
  - **CBRWorks**, da tec:Inno
  - **KATE**, da AcknoSoft
- A abordagem INRECA conjuga **Indução** e **CBR**:
  - A resolução de um problema inicia-se por recurso à árvore de decisão
  - Realizam-se *queries* à biblioteca de casos sempre que necessário, nomeadamente se existirem valores de alguns atributos que são desconhecidos no caso actual (ocorrência típica nos sistemas de diagnóstico)
  - A biblioteca de casos, recorrendo ao *k-nearest-neighbor*, devolve os **valores mais prováveis** para esses atributos, em face do conhecimento que tem de casos anteriores que apresentam com o actual uma elevada semelhança global.



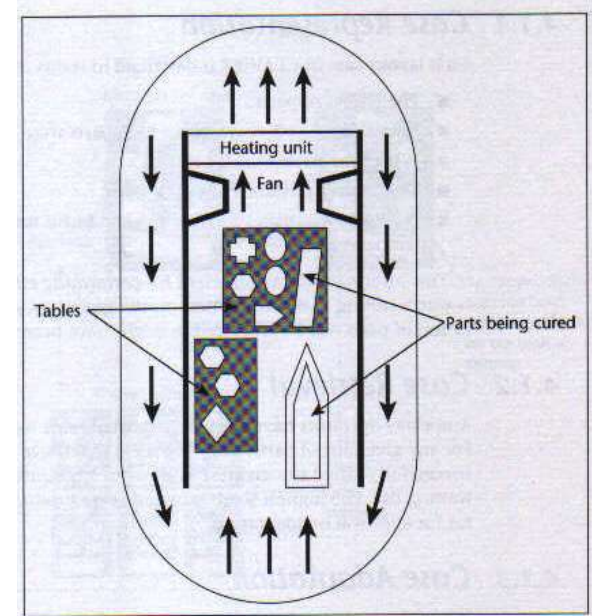
# 6.5 Protótipos com Aplicação Real

## 6.5 Protótipos com Validação de Campo

O mais significativo dos SP-CBR desta categoria é sem dúvida o conhecido

### CLAVIER (Mark, Simoudis & Hinkle, 1987)

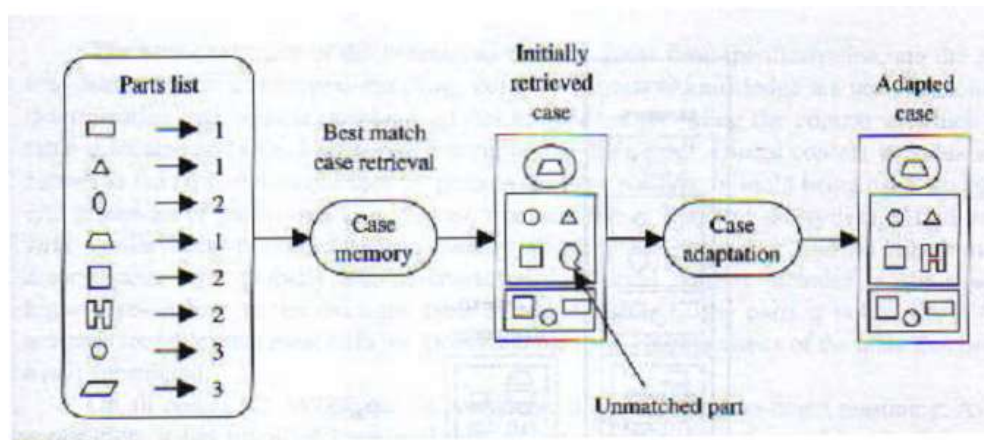
- Projecto iniciado em 1987 para a Lockheed Missiles and Space Company, U.S.A.
- Em uso regular desde 1990
- 1ª aplicação de campo do paradigma CBR
- Resultados óptimos, utilização noutras empresas do grupo e venda de direitos a outras empresas
- Planeia quais as peças de um material compósito que devem ser colocadas numa mesma carga para cura conjunta num autoclave.
- Planeamento necessário devido ao facto de o processo de aquecimento do autoclave não ser completamente compreendido e de a própria carga alterar o seu comportamento



*Autoclave e cura de peças no  
processo em que se usa o  
CLAVIER*

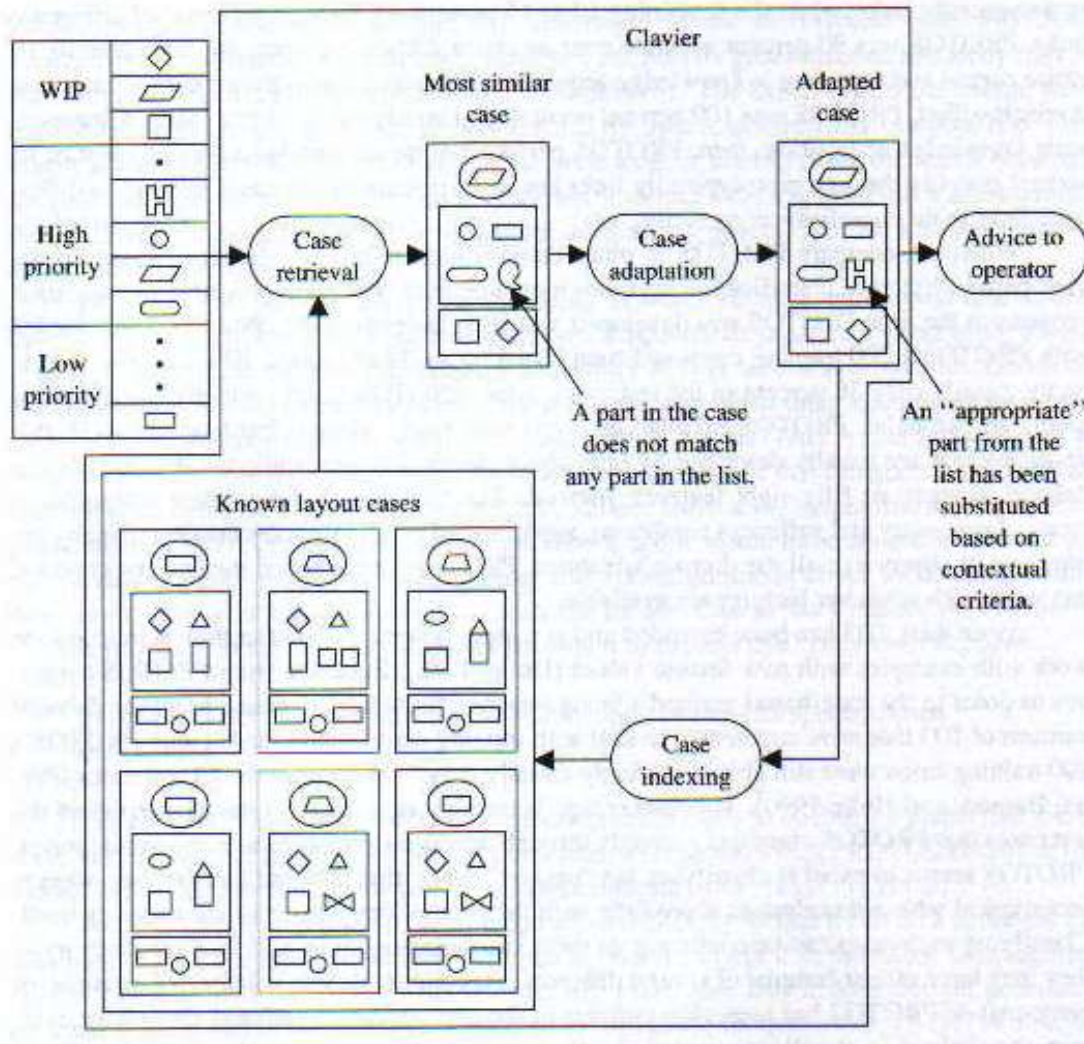
## 6.5 Protótipos com Aplicação Real

- Em função das **peças em lista de espera** e do **conhecimento acerca de cargas anteriores** configura as novas cargas



- Originalmente implementado em LISP num computador pessoal Macintosh
- Hoje encarado como memória colectiva da Lockheed
- Em 1994 o CLAVIER era utilizado noutras instalações da companhia, o licenciamento para outras indústrias estava em curso e os relatórios acerca de ciclos de fabrico mal sucedidos devido a cargas incompatíveis tinham sido virtualmente eliminados
- Anteriormente a Lockheed tinha tentado um SP baseado em regras, não tendo obtido resultados satisfatórios

# 6.5 Protótipos com Aplicação Real



Adaptação de  
casos no  
CLAVIER

NOTA:

Em (Leake et al., 1996 - editor) refere-se que os autores do CLAVIER constataram que os utilizadores preferem adaptar as cargas manualmente usando o sistema para verificar se houve problemas com cargas do tipo assim criado, no passado.



## 6.6 Shells-CBR

- MindBox *Art\*Enterprise*
- Halley Enterprise *The Easy Reasonner*
- SHAI *Esteem*
- I-Soft *Recall*
- empolis (tecIno) *kms-knowledge manager suite*
- Kaidara (Acknosoft) *Kaidara Advisor (KATE)*
- Attar Software *Minner (Indução de regras a partir de casos)*
- Inductive Solutions *Induce-IT*
- A Cognitive Systems (Remind) cessou em 1996



**The 5<sup>th</sup> International Conference on Case-Based Reasoning**  
**Trondheim, Norway, June 23 to 26, 2003**

**ICCBR 2003 is sponsored by:**



Kaidara e Empolis têm  
papéis de relevo no  
contexto CBR Europeu  
(e mundial)

<http://www.iccbr.org/iccbr03/>



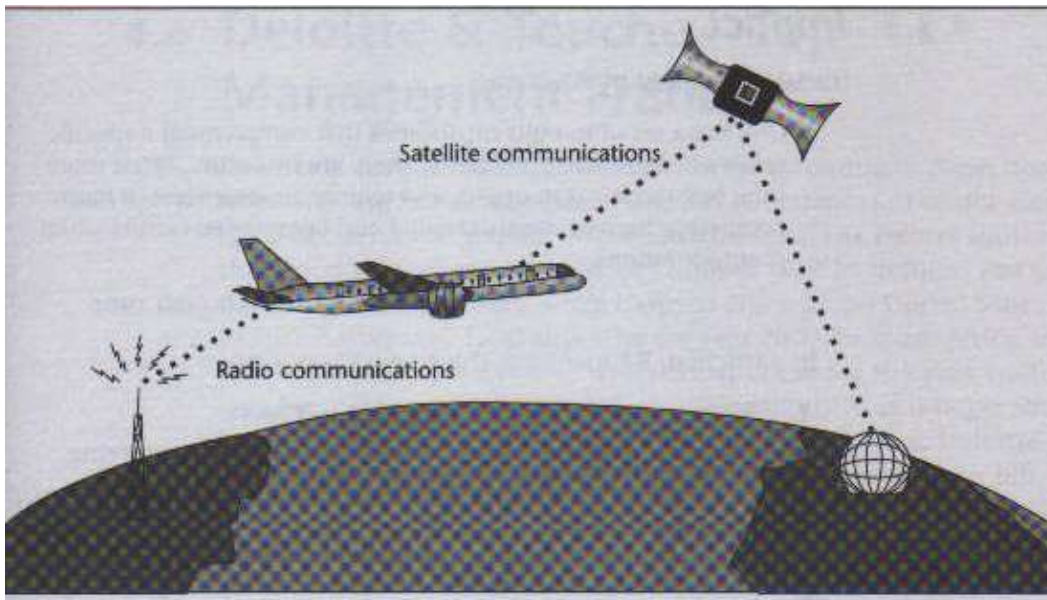
# 6.7 Aplicações e Casos de Estudo

## 6.7 Aplicações e Casos de Estudo

### 1. Remind

#### British Airways CASELine

- Diagnóstico de falhas em motores de Boeings 747-400
- O diagnóstico pode começar com o avião ainda em voo:



- Anomalia detectada a bordo (pilotos ou aparelhagem de monitorização) → Comunicação à equipa de manutenção em terra

# 6.7 Aplicações e Casos de Estudo

---

- **Operação:**

- Introdução de um número que corresponde a uma falha descrita no manual de manutenção do avião
- Texto preciso que designa uma falha
- Texto de conteúdo e comprimento variáveis
- A cada um destes métodos corresponde um sistema de pesquisa diferente, designado respectivamente por ATA, Mensagem EICAS ou Problema Relatado
- Podem ser usados isoladamente ou em cooperação
- Baseiam-se no nearest-neighbor ou na indução proporcionadas pelo Remind

- **Vantagens relatadas:**

- Redução significativa dos tempos de paragem:
- O custo de um avião em terra, é, em média, calculado em cerca de 1.500 dólares/minuto (ainda não contabilizando a necessidade de oferecer refeições ou estadia aos passageiros)
- É utilizado na British Airways, onde chegou a estar em consideração a sua aplicação à manutenção do Concorde (contudo este avião vai ser abandonado após o dramático acidente em Paris e consequente baixa do nº de passageiros)



# 6.7 Aplicações e Casos de Estudo

---

## 2. KAIDARA Advisor



### 2.1 CFM / BOEING: Manutenção de Turbinas CFM 56-3

CFM International: joint venture 50/50 da General Electric's e da Snecma.

Ao longo de 10 anos; 80 milhões de horas de voo; 23.000 relatórios de falhas do CFM 56-3 que equipam Boeings 737

- a) O tempo de diagnóstico é cerca de 50% do seu down-time
- b) O atraso numa partida pode custar à companhia \$1.500 por minuto
- c) A garantia de manutenção é parte integrante de qualquer estratégia de vendas

AcknoSoft (Kaidara) desenvolveu o CASSIOPÉE para diagnóstico de falhas: Biblioteca de casos a partir de relatos das companhias aéreas (inicial/ 16.000 cerca de 1.500, actualmente)

### Vantagens Relatadas:

- Redução do tempo de diagnóstico a metade (menos 25% no down-time)
- Redução dos atrasos nas partidas
- Redução dos erros de diagnóstico
- Base de know-how corporativo que a Snecma actualiza regularmente por CD-ROM (disponibilização e transferência de conhecimento dos peritos para jovens técnicos)
- O sistema interliga-se com um catálogo de suplentes e documentação electrónica, sendo utilizado por várias companhias aéreas em todo o mundo.

# 6.7 Aplicações e Casos de Estudo

---

## 2.2 ESTALEIROS DE ODENSE



### Manutenção de Robots de Soldadura

Estaleiros de Odense (*OSS - Odense Steel Shipyard*), Dinamarca: um dos líderes no projecto e construção de navios comerciais

Utilizam enormes robots de soldadura que soldam, cada um, cerca de 14m/h (com 25 destas máquinas, os OSS conseguem soldar mais de 2.5km por dia!)

Em 1997 lançaram um projecto sistema de diagnóstico do Gantry System, a estação robótica maior do mundo

Sistema resultante: Freya

- Desenvolvido em colaboração com várias empresas e inclui informação multimédia (documentação técnica, esquemas mecânicos e electrónicos, fotografias e filmes)
- Avisa os operadores acerca de qualquer falha num dos robots
- O Kaidara-Advisor pesquisa uma biblioteca de cerca de 500 casos

### Vantagens Relatadas:

- Down-time foi consideravelmente reduzido
- O sistema é utilizado na Alemanha, França, Finlândia e Noruega

# 6.7 Aplicações e Casos de Estudo

---

## 2.3 ARIANE

### Fiabilidade dos Lançamentos



Lançamento dos foguetões Ariane é responsabilidade do CNES - *Centre National d'Études Spatiales* - situado em Kourou, na Guiana Francesa.

Em 2002 foram realizados 12 lançamentos, esperando-se um número superior para 2003 (aproximadamente um em cada 3 semanas).

É fundamental uma elevada fiabilidade das instalações.

Para a garantir, o CNES apoiava-se numa base de dados que descrevia 10 anos de incidentes (cerca de 17.000, com 1.500 novos, todos os anos)

Em face de uma nova ocorrência estes dados eram analisados por controladores de qualidade que se socorriam de ferramentas estatísticas. Contudo, estas ferramentas eram incapazes de fornecer as soluções adoptadas em ocorrências semelhantes, no passado.

Baseando-se no Kaidara-Advisor, a MS&I - Matra Systems & Information - desenvolveu um sistema de diagnóstico capaz de substituir as pesquisa manuais realizadas sobre extensas listas de dados. O sistema é utilizado pelo Departamento de Qualidade e Manutenção de Kourou.



# 6.7 Aplicações e Casos de Estudo

---

## 2.4 NATIONAL SEMICONDUCTOR



### Controlo de Qualidade em Semiconductores

A National Semiconductor fabrica centenas de milhões de componentes electrónicos para milhares de clientes dispersos pelo mundo.

De entre estes componentes, cerca de 30 ppm - *parts per million* têm defeitos

A National mantém um grupo de trabalho para realização de PQAs - *Product Quality Analysis*

Desenvolveu *software* inspirado na metodologia 8D (*8-Discipline*) construído sobre uma base de dados relacional Sybase: não resolveu nunca o problema de modo satisfatório.

Os engenheiros, socorrendo-se de *queries* standard das RDBs, faziam uso de operadores lógicos e definiam gamas para os parâmetros intervenientes, mas obtinham geralmente *outputs* demasiado volumosos, ou nulos.

Paradigma CBR: "*Já ocorreu antes algo parecido com isto?*" -

**Ponderando a importância dos parâmetros** intervenientes na selecção, possibilitou a pesquisa eficiente da base de dados

O CBR passou a devolver casos semelhantes onde os *queries* clássicos devolviam "*no match found*"...

# 6.7 Aplicações e Casos de Estudo

---

## 2.4 Outros Case-Study Kaidara

- Manutenção do Metropolitano de Nápoles
- Diagnóstico de falhas em equipamento electrónico para helicópteros e aviões (nomeadamente Airbus A330 e A340) na Sextant;
- Diagnóstico de falhas na Freightliner, USA, fabricante de veículos pesados do grupo DaimlerChrysler.

## 3. Empolis Orenge / CBR-Works



### 3.1 SIEMMENS AG

#### Suporte on-line a Sistemas de Automação

A Siemens é líder de mercado no sector da automação industrial: 52.000 empregados em 66 unidades dispersas por todo o mundo, com facturação na ordem dos 7 biliões de Euros.

A partir da Exposição de Hannover em 1998, a Siemens disponibiliza na Internet o SIMATIC Knowledge Manager (SKM), um self-service de informações periciais

Fornece acesso a mais de 10.000 FAQs

A implementação deveu-se ao aumento de pedidos de assistência, catálogo muito diversificado, dependência de peritos e rápido aumento da documentação de suporte

# 6.7 Aplicações e Casos de Estudo

---

## 3.2 DEUTSCHE TELEKOM AG



### *Helpdesk* para Assistência Telefónica

A equipa de *helpdesk* da Deutsche Telekom (DT) situa-se em Elmshorn, Alemanha, e pode ser utilizado por técnicos, empresas e consumidores finais

Desde 1997 que o paradigma CBR tem sido utilizado como ajuda neste centro de assistência, tendo proporcionado, de acordo com a empresa:

- a) Disponibilização do conhecimento pericial;
- b) Percepção, por parte dos clientes, de um atendimento mais rápido e eficaz;
- c) Libertação dos peritos para tarefas mais complexas;
- d) Treino mais fácil e rápido das novas equipas técnicas.

Nas palavras da DT *"The system acts as a company-wide knowledge memory, supporting each individual person where desired"*.

Actualmente o centro presta assistência a 2000 problemas por dia, sobre 1800 produtos diferentes.

Já proporciona uma resolução imediata, pelo telefone, de 40% dos casos.

# 6.7 Aplicações e Casos de Estudo

## 3.3 Outros Case-Study Empolis

- Analog Devices
- Bayer
- Bosch
- Lufthansa
- Expo 2000 Hannover
- Fujitsu-Siemmens
- Hewlett Packard
- Kluwer Academic Publishers
- Shell
- Volkswagen
- etc..



Para mais informações, casos de estudo, demos de SPs e *downloads*, consultar

<http://www.empolis.com/>

Utilizaremos o CBR-Works nas Aulas Práticas

## 4. Stottler Henke AI (SHAI) - ESTEEM



### 4.1 Controlo de Produção na Indústria Siderúrgica

Uma aplicação interessante consiste na regulação dos set-points de linhas de produção de aço baseada em regulações utilizadas na produção destinada a encomendas anteriores.

# 6.8 CBR e Comércio Electrónico

---

## 6.8 CBR e Comércio Electrónico

A selecção de casos com base numa semelhança global e a possibilidade de ponderar os atributos tornam o sistema de selecção CBR óptimo para a pesquisa de produtos adequados às pretensões e perfil de um cliente.

Por isso, quase todas as empresas disponibilizam módulos especialmente vocacionados para o comércio electrónico **B2C - Business to Consumer** e **B2B - Business to Business** (Exemplo: Kaidara-commerce)

### HOOKE & McDONALD: Arrendamento (B2C)



Agência imobiliária líder na Irlanda

**e:km + lógica difusa** = procura de imóveis de arrendamento, on-line

Parâmetros de pesquisa: localização, tamanho, quartos, estacionamento, data...

As pesquisas retornam um match de 100% ou, na sua impossibilidade, o produto mais semelhante seguido de outros ordenados por semelhança decrescente.

- a) Aumento de sucesso nas pesquisas, actualmente de 54%;
- b) Resposta mais rápidas para os clientes;
- c) Facilidade de manutenção



## 6.8 CBR e Comércio Electrónico

---

### QUOKA AG: Anúncios Classificados - Automóveis (B2C)



Oferece mais de 1.500.000 de produtos on-line através de anúncios classificados. Encontra-se também em Portugal através da Ocasião, que publica o **jornal Ocasião**

Pesquisa de automóveis em 2ª mão:

- A especificação de características está sujeita a numerosas variantes:
  - Quantas formas há de descrever um carro? (incluindo o uso de abreviaturas). Veja-se o exemplo "Chrysler blu/pal, 4dr, 5sp, w/low mi trani, p4/w, AM/FM/cass, Xint. cond., 12,000/obo".
  - O sistema implementado valida as abreviaturas e permite que as consultas sejam realizadas em linguagem natural.

### ANALOG DEVICES: Selecção de Circuitos Integrados (B2B)



A Analog Devices (AD) é um dos líderes na produção de circuitos integrados utilizados em processadores de sinal digitais e analógicos.

Imprimia anualmente catálogos e data-sheets que perfaziam mais de 60cm de altura.

O custo, para 50.000 clientes dispersos pelo mundo, era de 3 milhões de dólares/ ano.

## 6.8 CBR e Comércio Electrónico

---

Suporte técnico por telefone:

- Engenheiros mantinham diálogo com os clientes enquanto tentavam seleccionar um produto com as características requeridas;
- Para lhes prestar apoio foram desenvolvidos 6 projectos que assentavam em bases de dados relacionais, mas cujos resultados eram listas muito extensas, ou "no match"

Solução Parametric Search (baseada no Kaidara-commerce)

- A AD distribui agora os seus catálogos em CD-ROMs (120.000) que incluem o Parametric Search. Este sistema permite-lhe economizar 2 milhões de dólares por ano e manter satisfeitos os clientes num mercado que é extremamente competitivo.

### **CHECK OUT TOURISTIK GMBH: Viagens de Última Hora**



- A Check Out Touristik GMBH opera em [www.reisenboerse.com](http://www.reisenboerse.com) um *site* destinado ao comércio de viagens de última hora
- Segundo a companhia, as propostas efectuadas por este sistema são iguais às que os técnicos da própria companhia fariam e os resultados disso são os melhores.
- Esta solução ganhou os prémios "Best of Internet World Berlin 99" e "Innovative Application of AI 1999", Orlando USA.

# 6.9 CBR e Ensino à Distância

---

## 6.9 CBR e Ensino à Distância

Pela sua natureza os sistemas baseados no paradigma CBR contêm exemplos reais que podem ser da máxima utilidade na exemplificação de matérias leccionadas por e-learning e mesmo no ensino tradicional

A integração entre plataformas de e-learning , CBR e ITS - *Intelligent Tutoring Systems* - pode dizer-se que ainda está em curso...



### F-16

Um dos produtos da Lockheed onde o CLAVIER trouxe o paradigma CBR ao mundo real ...