



PROSPECÇÃO E DESCOBERTA DE INFORMAÇÃO

Web Usage Mining

Luis Rafael - 2002



- Web Usage Mining -

- λ **Uso de técnicas de data mining para descobrir padrões de uso na web.**
- λ **Análise e descoberta de conhecimento através de dados e serviços web**
- λ **Descobrir e Extrair de forma automática informação dos documentos e serviços existentes na web, conh. novo, relevante**



- Web Usage Mining -

λ Tipos/categorias de Web Mining

- Web content mining

(descoberta de info. útil a partir do conteúdo/dados/documentos da web)

- Web structure mining

(descoberta de modelos dentro da página: ligações, html xml etc.)

- Web usage mining

(descoberta de informação com base nos dados gerados a partir das sessões de utilização na web: ip, tempo, produtos...)

- User profile

(informação demográfica dos utilizadores: registo)



- Web Usage Mining -

λ Fontes dos dados

- **Server level collections** (acesso a serv. Web gera o registo das pág. visitadas)
- **Client level collections** (seguimento do que faz utiliz. Sites visitados: cookies, sniffers)
- **Proxy level collection** (intermediário/procurador entre browser do cliente e serv. Web minimizar tempo/ /carga rede armazena em cache páginas acedidas por proxy, revela comportamentos de grupos de util. anónimos)



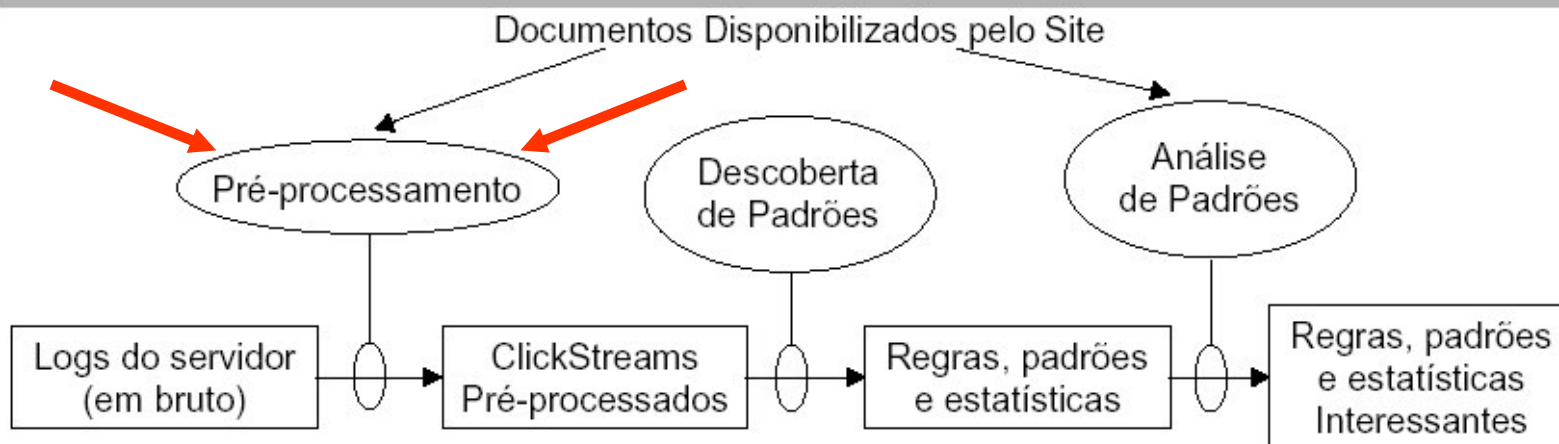
- Web Usage Mining -

λ O que faz o Web Usage Mining

- Pré-processamento
- Descoberta de padrões
- Análise de padrões

- Web Usage Mining -

Uma visão global do processo de Web Usage Mining





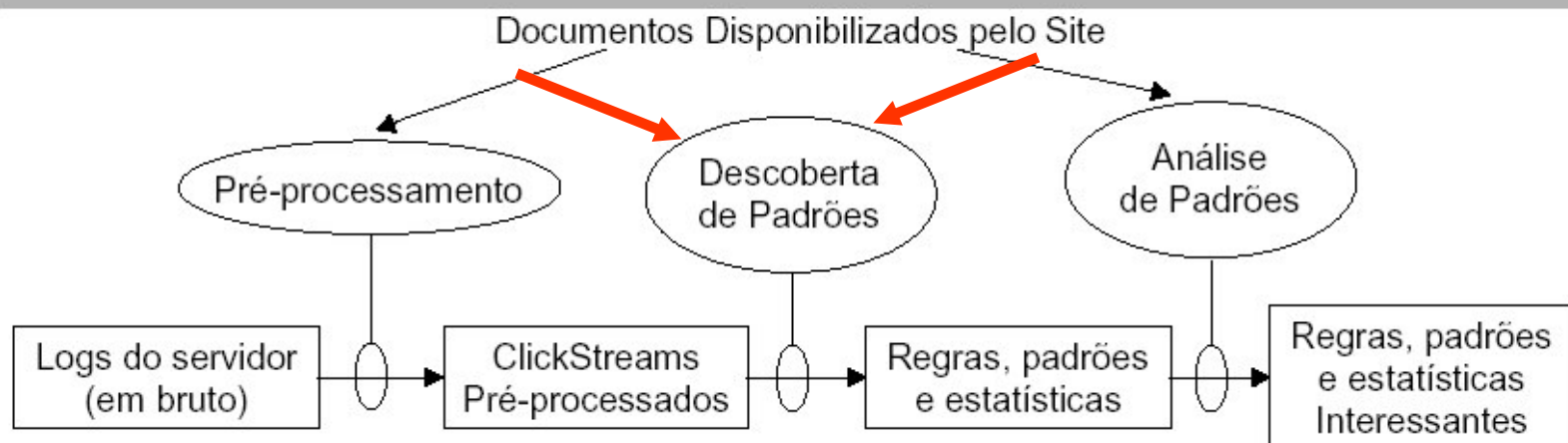
- Web Usage Mining -

λ Pré-processamento

- λ Tratamento/conversão do conteúdo, estrutura, dos dados disponíveis para descoberta de padrões
 - Pré-tratamento (limpeza)
 - Pré-processamento do conteúdo (converter texto e imagens, scripts em dados usáveis para a Desc. Pad.)
 - Pré-processamento da estrutura (links)

- Web Usage Mining -

Uma visão global do processo de Web Usage Mining





- Web Usage Mining -

λ Descoberta de Padrões

- λ (descoberta automática de padrões em websites individuais ou sites múltiplos)
- λ (extração de padrões usando métodos e algoritmos para identificar, classificar e agrupar)
 - Análise estatística
 - Regras de associação
 - Clustering
 - Padrões sequenciais
 - modelo de dependência
 - Aprend. Aut.

- Web Usage Mining -

Uma visão global do processo de Web Usage Mining





- Web Usage Mining -

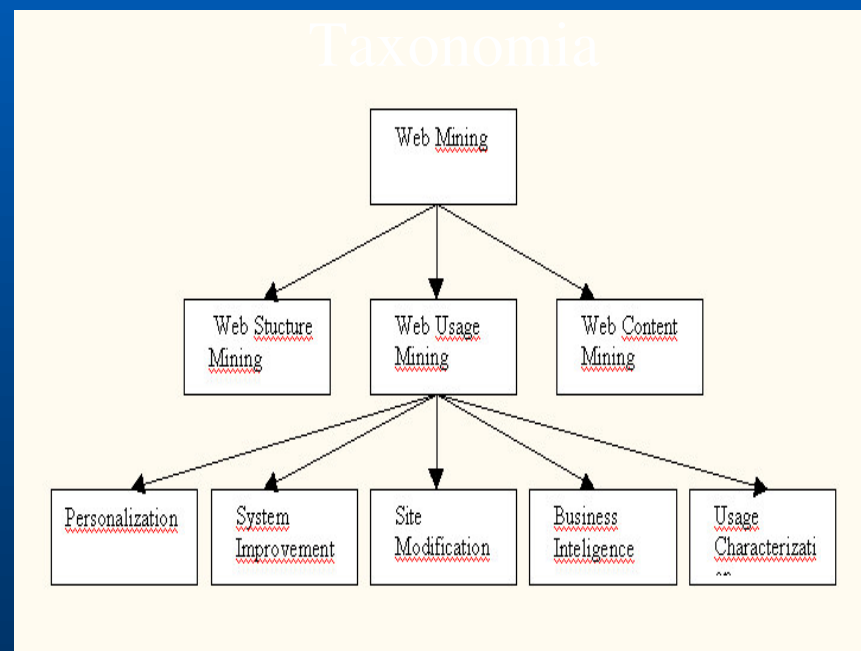
λ Análise de padrões

- λ (validação e/ou interpretação dos padrões descobertos)
- λ Saída de regras interessantes ou não
- λ Padrões de conjuntos encontrados na Desc.Pad.
- λ Análise de padrões: SQL, OLAP...
 - Problema: quant. de regras, difícil, avaliar e interpretar como conhecimento útil e relevante

- Web Usage Mining -

λ Project survey (Especializações)

- Personalização()
- melhorias do sistema
- modificações no site
- inteligência negocial (suporte decisão. em e-com)
- caracterização do USO (usabilidade qualidade)





- Web Usage Mining -

- λ **Personalização web baseada no uso**
 - **Off-line:** a) pré-processamento, prep. dados limpeza, filtragem ...
b) conhecer e descobrir padrões de uso(regras ass...)

 - **On-line: Recomendações**
Sugestões p/sessão corrente (páginas que irá(?) querer visitar etc...)



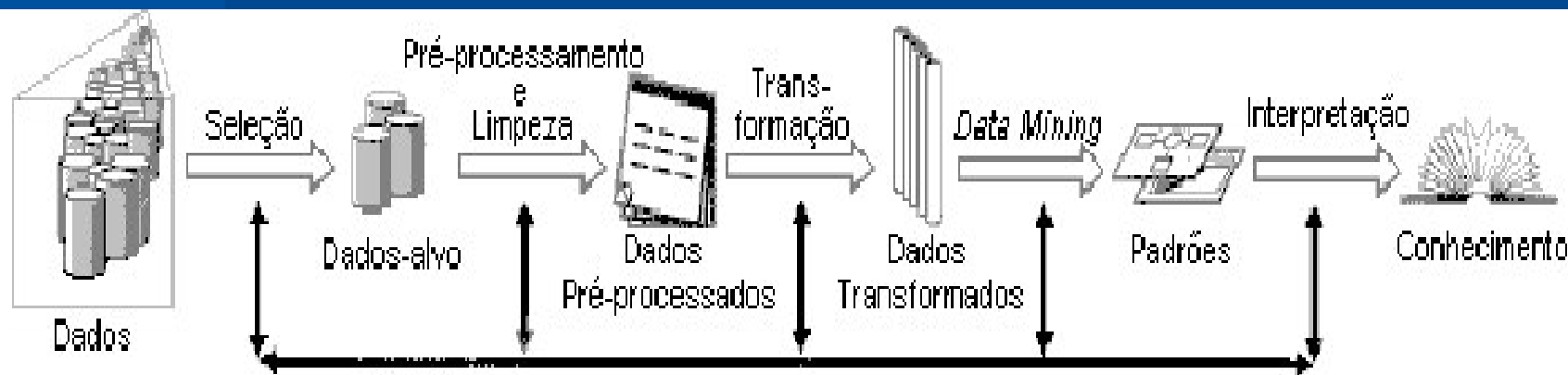
- Web Usage Mining -

λ Conclusão

- Aplicar o conhecimento adquirido para melhor servir o cliente (ex.comércio electrónico)
- Oportunidade para personalizar as exp. dos util./cliente.
- Aprendizagem automática preferências
- Melhorar a eficiência e o sucesso
- Web usage mining vai buscar técnicas a outras áreas: KDD, IA,...
- A descoberta e análise de padrões de navegação de utilizadores na web, está baseado nos conceitos do processo KDD*

- Web Usage Mining -

Uma visão global do processo de Web Usage Mining



Passos do processo de KDD.