

# Linguagens Documentárias

Profa. Lillian Alvares

Faculdade de Ciência da Informação, Universidade de Brasília

# Contexto

---

- ▶ Organização da Informação...

- ▶ ... procura criar **métodos e instrumentos** para elaborar informação documentária...

- ... Indexar, resumir e construir linguagens de representação são os termos técnicos que denominam essas operações



# Contexto

---

- ▶ A questão é:
  - ▶ Como organizar a informação para que o conhecimento fique disponível?



# Contexto

---

- ▶ Pelas Linguagens Documentárias, pois...
  - ▶ ...elas estão contidas no contextos de produção e de busca de informação...
  - ... é parte constitutiva da preocupação com a recuperação da informação.



# Contexto

---

- ▶ Uma das condições para plena recuperação de informação é o uso de Linguagens Documentárias



# Contexto

---

- ▶ As Linguagens Documentárias são instrumentos privilegiados de mediação que apresentam **dupla função**:



# Contexto

---

- ▶ a) **representar** o conhecimento
- ▶ b) promover **interação entre usuário e conteúdo.**



# Principais intersecções interdisciplinares

---

- ▶ **Linguística:** o princípio estrutural da organização da linguagem
- ▶ **Terminologia:** a modelagem do conceito e dos sistemas de conceitos, onde, glossários e dicionários terminológicos especializados são seus produtos
- ▶ **Lógica:** a identificação das formas de raciocínio e de organização de conjuntos



# Campos para atuação...

---

- ▶ **Não é exclusivo** dos Bibliotecários e Documentalistas;
- ▶ Áreas de aplicação: Filosofia, Informática, Inteligência Artificial, Engenharia, Psicologia entre outros....



# Contexto

---

- ▶ As LD constituem uma espécie de **código de tradução** que tem, entre suas funções, a normalização das representações documentárias como meio de **viabilizar sua comunicação**.



# Contexto

---

- ▶ Essa tradução que permite a **transformação de informações dos textos em linguagem natural e/ou especializada para uma linguagem documentária controlada**, envolve a questão da significação:
  - ▶ Como representar adequadamente as informações **sem comprometer o seu significado?**



# Resultados da tradução

# Contexto

---

- ▶ LD são instrumentos **que não permitem a recuperação da dinamicidade dos textos**
- ▶ um dos seus objetivos é o de fornecer uma **versão extremamente econômica** dos textos originais
  - ▶ por intermédio, do **destaque das informações** consideradas **centrais** nos documentos.



# Contexto

---

- ▶ LD e textos são construções de natureza distinta, por isso..
  - ▶ ... na **operação de conversão perde-se a especificidade do texto original em prol da generalização.**





# Definições

# Definições

---

- ▶ São linguagens **artificialmente construídas** que visam “**traduzir**” sinteticamente **conteúdos documentais**, utilizadas nos sistemas documentários para **indexação, armazenamento e recuperação da informação**.



# Definições

---

- ▶ É um **conjunto controlado de termos** usado para representar os **assuntos dos documentos**.



# Definições

---

- ▶ O objetivo da LD é o **controle de vocabulário**, isto é, **controlar a terminologia** de área ou áreas do conhecimento por meio do estabelecimento de um conceito/interpretação definido aos termos de acordo com as necessidades de uso do sistema.



# Definições

---

- ▶ A utilização de uma LD para escolha do termo correto para descritor **reduz a diversidade e a ambiguidade da terminologia** e estabelece a eficácia com que o indexador pode descrever o assunto dos documentos, ***já que vários autores podem se utilizar de diferentes palavras para expressar uma mesma idéia.***



# Definições

---

- ▶ Só podem cobrir conceitos de um domínio específico do conhecimento humano, não havendo assim, linguagens documentárias gerais.



# Definições

---

- ▶ Visam a uniformização do uso de linguagens especializadas...
  - ▶ ...proporcionando uma representação padronizada do conteúdo informacional...
    - ... **bem como uma recuperação da informação mais pertinente.**



# Definições

---

- ▶ As palavras (que descrevem as informações ou descritores) são os elementos que compõem as linguagens documentárias.
  - ▶ **São extraídos da linguagem natural**



# Terminologia

---

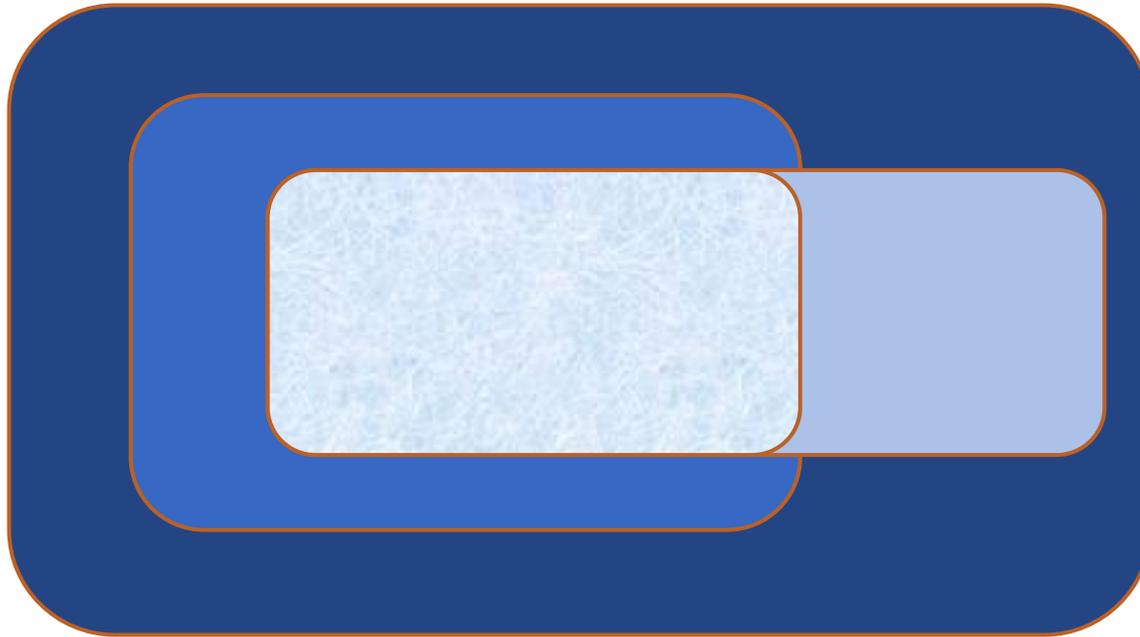
- ▶ Tipos de linguagem
  - ▶ Linguagem Natural
  - ▶ Linguagem Documentária, também conhecida como:
    - **Linguagem Controlada**
    - **Linguagem de Indexação**
    - **Linguagem Descritoras**
    - **Codificações Documentárias**



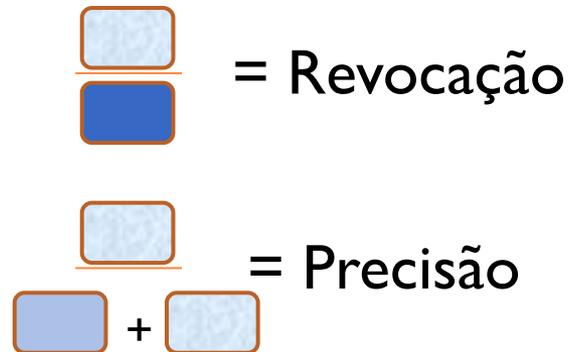
# Medidas de Eficácia

# Revocação e Precisão

---



	<b>Todos os documentos</b>
	Documentos relevantes
	Documentos retornados
	Relevantes retornados



# Revocação ou Recall

---

- ▶ Relação entre o número de documentos relevantes sobre determinado tema, recuperados pelo sistema, e o número total de documentos sobre o tema, existentes nos registros do mesmo sistema.

# Precisão ou relevância

---

- ▶ Refere-se à qualidade dos documentos recuperados pelo sistema, para atendimento dos pedidos relativos a temas específicos, encaminhados por usuários diversos.
- ▶ Documentos relevantes retornados dividido pelo número total de retornados



# Linguagens Pré-Coordenadas e Linguagens Pós-Coordenadas



# Linguagens Pré-Coordenadas

# Linguagens pré-coordenadas

---

- ▶ **Combinação feita na etapa de entrada das palavras**, na indexação.
- ▶ Os assuntos são representados por um conjunto de termos já combinados.
  - ▶ Exemplo: Álbum de fotografia



# Linguagens pré-coordenadas

---

- ▶ Devem contemplar todas as possibilidades de combinação entre os termos para formar assuntos complexos.
  - ▶ São palavras, termos ou frases escolhidas para expressar um conceito ou uma combinação de conceitos de indexação



# Linguagens pré-coordenadas

---

- ▶ Adota:
  - ▶ VER: remete ao termo adotado como entrada (sinônimos, termos específicos para termos genéricos)
  - ▶ VER TAMBÉM: remete a assuntos relacionados.  
Complementa a informação para o usuário



# Linguagens pré-coordenadas

---

- ▶ Trazem problemas à indexação, devido, principalmente, impossibilidade de prever todas as alternativas de relacionamento entre conceitos.



# Linguagens pré-coordenadas

---

- ▶ O número e a precisão dos termos de uma lista podem levar a crer numa busca facilitada.
- ▶ Não são flexíveis para permitir uma busca por qualquer termo que compõe o cabeçalho.



# Linguagens pré-coordenadas

---

- ▶ Podem ser de dois tipos:
  - ▶ Classificatórias (Classificações)
  - ▶ Alfabéticas (Cabeçalhos de assunto)



# Desvantagens

---

- ▶ Necessitam de entradas múltiplas para explicitar todos os conceitos significativos
- ▶ Difíceis de atualizar, pouco dinamismo
- ▶ Menor flexibilidade na indexação
- ▶ Relações estáticas
- ▶ Menor revocação





# Linguagens Pós-Coordenadas

# Linguagens pós-coordenadas

---

- ▶ Combinação feita no momento da saída, na recuperação da informação.
- ▶ Utilização de operadores booleanos: E, OU, NÃO
  - Exemplo: álbum E fotografia para recuperar “álbum de fotografia”



# Vantagens

---

- ▶ Dinâmicas, novos termos podem ser incluídos com maior rapidez
- ▶ Relações multidimensionais entre termos
- ▶ Todos os termos são pontos de acesso
- ▶ Maior flexibilidade na indexação



# Desvantagens

---

- ▶ Geram falsa recuperação, necessitam de artifícios para evitá-la
- ▶ Exigem maior esforço na busca





# Linguagem Natural (LN) e Linguagem Documentária (LD)

LN	LD
A parte <u>oral</u> é mais importante.	São fundamentalmente <u>escritas</u> .
São <u>gerais</u>	São <u>especializadas</u>
São estabelecidas e adaptadas através de <u>longos períodos</u> de tempo e por muitas pessoas.	São estabelecidas em <u>poucos anos</u> por um número reduzido de pessoas
A <u>sinonímia e a polissemia</u> são fatos naturais das LN. Aceitam ambigüidades	<u>Não aceitam ambigüidades</u> . Não comportam polissemias e se esforçam por evitar as sinonímias.
São <u>naturalmente aceitas</u> e <u>adquiridas</u> pelos usuários	<u>Devem ser aceitas</u> pelos usuários.
Interessam os aspectos <u>fonológicos, semânticos e sintáticos</u>	Interessam somente os aspectos <u>semânticos e sintáticos</u>
São <u>naturais</u>	São <u>artificiais</u>

LN	LD
São <b><u>menos eficientes</u></b> que as LD nas operações de recuperação da informação	São <b><u>mais eficientes</u></b> que as LN nas operações de recuperação da informação
São <b><u>instrumentos</u></b> de <b><u>comunicação</u></b>	São <b><u>instrumentos</u></b> de <b><u>comunicação</u></b>
Têm sua <b><u>própria estrutura</u></b>	Sua estrutura <b><u>baseia-se na estrutura da LN</u></b> sobre a qual elas são formadas
Necessitam de <b><u>hierarquias para evitar</u></b> malformações gramaticais	Necessitam de <b><u>hierarquias para evitar</u></b> malformações gramaticais.
São <b><u>sensíveis a mudanças culturais</u></b>	São <b><u>sensíveis a mudanças culturais</u></b>
<b><u>Não têm uma função específica</u></b>	<b><u>Têm um propósito específico e</u></b> um nível de funções



# LD: Vantagens e desvantagens

---

VANTAGENS	DESVANTAGENS
<p>Se bem construído, o vocabulário controlado poderá oferecer <b><u>alta precisão e revocação</u></b>, e também <b>ampliar a confiança do usuário</b> frente a um possível resultado negativo</p>	<p>Um rígido vocabulário controlado <b>podará não refletir os objetivos</b> da base de dados</p>
<p>As relações hierárquicas e as remissivas do vocabulário controlado <b>auxiliam tanto o indexador, quanto o usuário</b> na identificação de conceitos relacionados</p>	<p>Necessidade de <b>treinamento</b> no uso dos vocabulários controlados tanto para os intermediários, quanto para os usuários finais</p>



# LD: Vantagens e desvantagens

---

<b>VANTAGENS</b>	<b>DESVANTAGENS</b>
<p><b>Redução no tempo de consulta</b> à base, pois a <b>estratégia de busca</b> será <b>melhor elaborada</b> com o uso do vocabulário controlado</p>	<p><b>Desatualização</b> do vocabulário poderá conduzir a resultados falsos</p>
<p><b>Aumenta a precisão</b></p>	



# LD: Vantagens e desvantagens

---

<b>VANTAGENS</b>	<b>DESVANTAGENS</b>
<p><b>Controle total do vocabulário</b> de indexação, minimizando os problemas de comunicação entre os indexadores e os usuários</p>	<p><b>Custos:</b> a produção e manutenção da base de dados terão despesas maiores com a equipe de indexadores. Terá ainda que, <b>manter pessoal especializado</b> na atualização do vocabulário</p>
<p>Promove a <b>consistência/coerência na indexação</b></p>	





# Linguagem Natural

# LN: Vantagens e desvantagens

---

VANTAGENS	DESVANTAGENS
Permite o <b>imediato registro</b> da informação numa base de dados, sem necessidade de consulta a uma linguagem de controle	Os usuários, no processo de busca, precisam fazer um <b>esforço intelectual maior para determinar os sinônimos, as grafias</b>
Processo de <b>busca é facilitado com a ausência de treinamentos específicos</b> no uso da linguagem de controle	Haverá uma <b>alta incidência</b> de <b>respostas negativas</b> ou de relações entre os termos usados na busca (por ausência de padronização)
Termos de <b>indexação são extraídos diretamente dos documentos que vão compor a base de dados</b>	<b>Custos de acesso tendem a aumentar com</b> a entrada de <b>termos aleatórios</b>



# LN: Vantagens e desvantagens

---

<b>VANTAGENS</b>	<b>DESVANTAGENS</b>
Termos <b>específicos citados</b> nos documentos <b>podem ser encontrados</b>	Uma <b>estratégia de busca que arrole todos os principais</b> conceitos e seus sinônimos deve ser elaborada para cada base de dados
<b>Elimina os conflitos de comunicação entre os indexadores e os usuários</b> , pois, ambos terão acesso aos mesmos termos	<b>Perda de confiança do usuário numa possível resposta negativa</b>





# Tipos de LD

# Tipos

---

## Pré-coordenadas

- ▶ Sistemas de classificação
- ▶ Listas de cabeçalhos de assuntos
- ▶ Glossários
- ▶ Dicionários

## Pós-coordenadas

- ▶ Vocabulários controlados
- ▶ Tesouro
- ▶ Taxonomias
- ▶ Ontologias



# Referência

---

- ▶ KOBASHI, Nair Yumiko. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. DataGramZero: Revista de Ciência da Informação, v.8, n.6, dez, 2007.

