

Vocabulários Controlados

baseado em:

ANSI/NISO (2005), *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*

- 1 Vocabulários controlados
- 2 Tipos de vocabulários controlados
- 3 Construção de um tesauro

Todo instrumento traz
consigo o espírito de sua
criação.

W. Heisenberg

- 1 Vocabulários controlados
- 2 Tipos de vocabulários controlados
- 3 Construção de um tesauro

Definições

- Vocabulários baseados em listas lexicais
- A necessidade dos vocabulários controlados surge de fatores intrínsecos à experiência linguística:
 - Divergências na interpretação do léxico (variações conceptuais)
 - Divergências no uso de expressões lexicais (variações sociais)
 - Ampliação da significação do léxico (polissemia)
 - Desconhecimento do léxico
- A organização e a sofisticação do vocabulário dependem do uso a que se destinam

Usos linguísticos

- Usos “linguísticos”:
 - Dicionários de analogias
 - Dicionários de sinônimos (bilíngues ou monolíngues)
- Usos “computacionais”:
 - Listagens do léxico disponível em um determinado contexto (ex.: *tag clouds*)
 - Listas de homógrafos (desambiguação)
 - Listas de homônimos (desambiguação)
 - Verificação da entrada de dados do usuário

Objetivos computacionais

Diminuição da divergência semântica entre diferentes sujeitos

- De todos os potenciais problemas enfrentados nos usos computacionais, o mais sério é o da **sinonímia**
- É preciso entender as razões de surgimento dos sinônimos nos discursos e os meios de contorná-las
 - Contextos de uso
 - Hábitos linguísticos (formas preferidas)
 - Desconhecimento de alternativas
 - Efeitos retóricos

Propostas de solução

Induzir à escolha de uma forma preferencial

- Acréscimo de informação lexical: apresentar ao usuário variantes desconhecidas ou pouco usadas por ele
- Acréscimo de informações sobre o léxico:
 - Apresentar contextos e exemplos
 - Apresentar estatísticas de uso ou ordenações de formas preferenciais

- 1 Vocabulários controlados
- 2 Tipos de vocabulários controlados
- 3 Construção de um tesauro

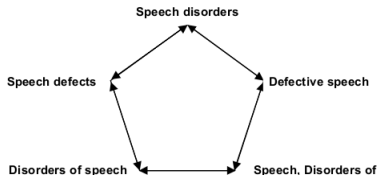
Tipos de vocabulários controlados

- Listas
- Redes de sinônimos
- Taxonomias
- Tesauros

Listas

- Simples repertórios de entradas lexicais
- Geralmente indexadas de acordo com algum critério (ex.: ordem alfabética) a fim de facilitar a consulta dos usuários ou as consultas dos Bancos de Dados
- Ex.: lista de UFs nos sites que pedem informações de endereço
- O critério de seleção das entradas que vão compor a lista é semântico, mas não há ordenação semântica dos dados na própria lista
- Controle vocabular oferecido: ambiguidade.

Redes de sinônimos

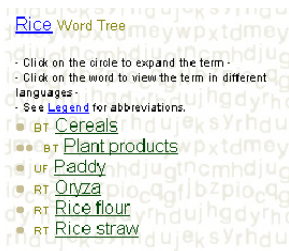


- Em inglês: *synonymy rings*
- As entradas são sinônimas umas das outras
- A disposição gráfica marca a afinidade nocional das entradas. Ao contrário do que o nome indica, os sinônimos podem ser dispostos em forma de lista, desde que seja mantido o critério semântico de ordenação dos termos
- Controles vocabulares: ambiguidade; sinonímia.

Taxonomias

- Organização hierárquica dos dados
- Funciona como dicionário definicional ao fornecer relações de hiponímia, hiperonímia e co-hiponímia
- Controles vocabulares: ambiguidade; sinonímia; relações hierárquicas

Tesouros



- Organização associativa dos dados
- Pode (deve) oferecer dicas de uso (UF, *used for*) e estatísticas
- As relações associativas são recíprocas e por isso devem ser redundantemente indicadas (através de abreviações convencionais e legendadas)
- Controles vocabulares: ambiguidade; sinonímia; relações hierárquicas; relações associativas

- 1 Vocabulários controlados
- 2 Tipos de vocabulários controlados
- 3 Construção de um tesauro

Considerações iniciais

- Como em qualquer problema computacional, é preciso saber a razão de ser do tesauro: isso determina o tamanho da massa de dados, o nível de detalhamento da hierarquia e a forma de apresentação do léxico
- O mais comum é que o vocabulário seja específico, isto é, ligado a uma área (aviação, por ex.) ou instituição
- Quais serão as fontes de informações do tesauro? Textos, dicionários, documentos internos da organização contratante. . .

Formato dos dados

- Perspectiva extensional: preferência por *substantivos* e no *plural*
- Formas preferidas devem vir marcadas, por ex., com (P)
- É útil dispor os dados em uma planilha
- A identificação por números das entradas é opcional

Formato dos dados

Formas compostas São encorajadas, porque aumentam a especificidade semântica da entrada. Ex.: Chave inglesa, Alicate de pressão.

Pré-coordenação É melhor apresentá-las sempre que possível, ao invés de deixar a possibilidade de pós-coordenação em aberto. Exs.: Aquecedores de água, Aquecedores de ambiente.

Exercício

- ① Construa um tesauro sobre animais
- ② Inclua no tesauro a entrada “Macaco-prego”