

# METADADOS E TAXONOMIAS

Organização da Informação

# Organização de Informação

- Será que existe uma necessidade básica de colocar coisas em ordem?
  - Muito da linguagem natural é sobre **categorias** de coisas ao invés de coisas individuais
  - Por que organizamos coisas e informação?
    - Por que as colheres vão em uma gaveta na cozinha e não em uma caixa no quarto?
    - Por que seus livros favoritos vão em uma estante e outros em outra?

# Por que organizar informação?

- Para encontrá-las mais facilmente!
- Recuperação eficiente está relacionada com algum tipo de organização
- Historicamente, existiram muitas instituições e ferramentas para organização de informação
  - Bibliotecas
  - Museus
  - Arquivos
  - Índices e catálogos, dicionários, cadernos telefônicos, etc.

# Por que organizar informação?

## □ Questão de escala

- Usar suas próprias categorias e métodos (ad hoc) para organizar sua coleção de livros, fotos ou CDs parece funcionar bem...
- Mas e se a sua coleção aumentasse para...
  - 10 vezes o tamanho?
  - 100 vezes?
  - 1000 vezes?
  - 100000 vezes?
- Será que a organização seria a mesma?

# O que é organização de informação?

- Identificar a existência de tipos de entidades atreladas à informação à medida que fiquem disponíveis
- Sistemáticamente agrupar estas entidades em coleções em bibliotecas, arquivos, museus, arquivos na Internet e outros repositórios

# O que é organização de informação?

- Produção de listas de entidades de acordo com regras padronizadas
- Fornecer nome, títulos, abstract e outras formas úteis de acesso a estas entidades
- Fornecer meios para localizar cada entidade ou uma cópia dela

# Questões chave

- Como descrever recursos de informação de forma que possam ser utilizados por quem precise?
  - ▢ *Organizar*
- Como encontrar os recursos de informação ou objetos apropriados para dadas necessidades?
  - ▢ *Recuperar*

# Na Gestão do Conhecimento...

---

- Conhecimento envolve um conjunto *organizado* de informações, que você *entende* e está familiarizado com o seu uso



# Compartilhamento de Conhecimento

- É fundamental:
  - Entendimento comum
  - Organização deste conhecimento
  - Disponibilidade (fácil acesso e disseminação)

# Gerência de Informação

- É fundamental:
  - Descrição consistente de recursos
  - Entendimento correto do que são e de como podem ser utilizados

# Mecanismos fundamentais

- Metadados - descrição de recursos
- Vocabulário comum
- Mecanismos de Classificação
- Representação de conceitos
- Compartilhamento de conceitos
- Inferência para obtenção de novo conhecimento
- ...



# METADADOS

# Metadados

## □ Meta

- do Grego: “sobre“, “além”
- Wikipedia: “It is a common English prefix, used to indicate a concept which is an abstraction from another concept, used to analyze the latter”
- e.g. metaphysics, meta-object, meta-information, Metadados
- “algo que descreve”

## □ Metadados

- Mais do que “dados sobre os dados”

*Quaisquer informações que permitam identificar, localizar, utilizar e entender os dados ou qualquer outro tipo de acervo*

# Metadados

- *Estruturas e linguagens para a descrição de recursos de informação e seus elementos (componentes ou características)*
- “Metadados são informações sobre a organização dos dados, os vários domínios de dados, e o relacionamento entre eles” (Baeza-Yates p. 142)

# Metadados

- Frequentemente distinguimos 2 principais tipos de metadados
  - **Descritivos**
    - Descreve o objeto de informação/dados e suas propriedades
    - Pode usar uma variedade de formatos e regras descritivas
  - **Tópicos**
    - Descreve o tópico ou “sobre que” de um objeto de informação/dados
    - Pode incluir uma variedade de vocabulários para descrição de assuntos, tópicos, categorias, etc.

# Sistemas e padrões de metadados

- Sistemas de nomeação e ID
- Descrição bibliográfica
  - Textos
- Musica
- Imagens e objetos
- Dados numéricos
- Dados geoespaciais
- Coleções
- Vídeo e filmes



# Mesmo item, Sistemas diferentes

---

- ISBD
- TEI Header
- MARC Record
- Dublin Core

# ISBD (Registro bibliográfico)

Introduction to cataloging and classification / Bohdan S. Wynar. -- 8th ed. / Arlene G. Taylor. -- Englewood, Colo. : Libraries Unlimited, 1992. -- (Library science text series).

[https://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons\\_2007-en.pdf](https://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_2007-en.pdf)

# MARC Record

ID:DCLC9124851-B RTYP:c ST:p FRN: MS:c EL: AD:06-20-91  
CC:9110 BLT:am DCF:a CSC: MOD: SNR: ATC: UD:04-11-92  
CP:cou L:eng INT: GPC: BIO: FIC:0 CON:b  
PC:s PD:1992/ REP: CPI:0 FSI:0 ILC:a II:1  
MMD: OR: POL: DM: RR: COL: EML: GEN: BSE:

010 9124851

020 0872878112 (cloth)

020 0872879674 (paper)

040 DLC\$cDLC\$dDLC

050 00 Z693\$b.W94 1991

082 00 025.3\$220

100 1 Wymar, Bohdan S.

245 10 Introduction to cataloging and classification /\$cBohdan S. Wymar.

250 8th ed. /\$bArlene G. Taylor.

260 Englewood, Colo. :\$bLibraries Unlimited,\$c1992.

300 xvii, 633 p. :\$bill. ;\$c24 cm.

440 0 Library science text series

504 Includes bibliographical references (p. 591-599) and index.

650 0 Cataloging.

650 0 Subject cataloging.

650 0 Classification\$xBooks.

630 00 Anglo-American cataloguing rules.

700 10 Taylor, Arlene G.,\$d1941-

# TEI Header

<https://tei-c.org/>

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title> Introduction to cataloging and classification</title>
      <respStmt><name>Bohdan S. Wynar<resp> 8th edition by</resp>
        <name>Arlene G. Taylor</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <distributor>Libraries Unlimited</distributor>
    </publicationStmt>
    <sourceDesc>
      <bibl> Introduction to cataloging and classification / Bohdan S. Wynar. -- 8th ed. / Arlene G.
      Taylor. -- Englewood, Colo. : Libraries Unlimited, 1992.
      </bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiHeader.html>

# Dublin Core

- Metadados simples para descrever recursos na internet
- Para “objetos-documento”
- 15 elementos

<https://dublincore.org/>

<https://dublincore.org/specifications/dublin-core/>

# Dublin Core

**TITLE:** Introduction to cataloging and classification

**CREATOR:** Taylor, Arlene G.

**OTHER CONTRIBUTOR:** Wynar, Bohdan S.

**DATE:** 1992

**FORMAT:** BOOK

**LANGUAGE:** ENG

**PAGES:** 633

**PUBLISHER:** Libraries Unlimited

**SUBJECT:** Cataloging.

**SUBJECT:** subject cataloging.

**SUBJECT:** Classification -- Books

**DESCRIPTION:** Textbook on cataloging and classification

**RESOURCE TYPE:** text.monograph

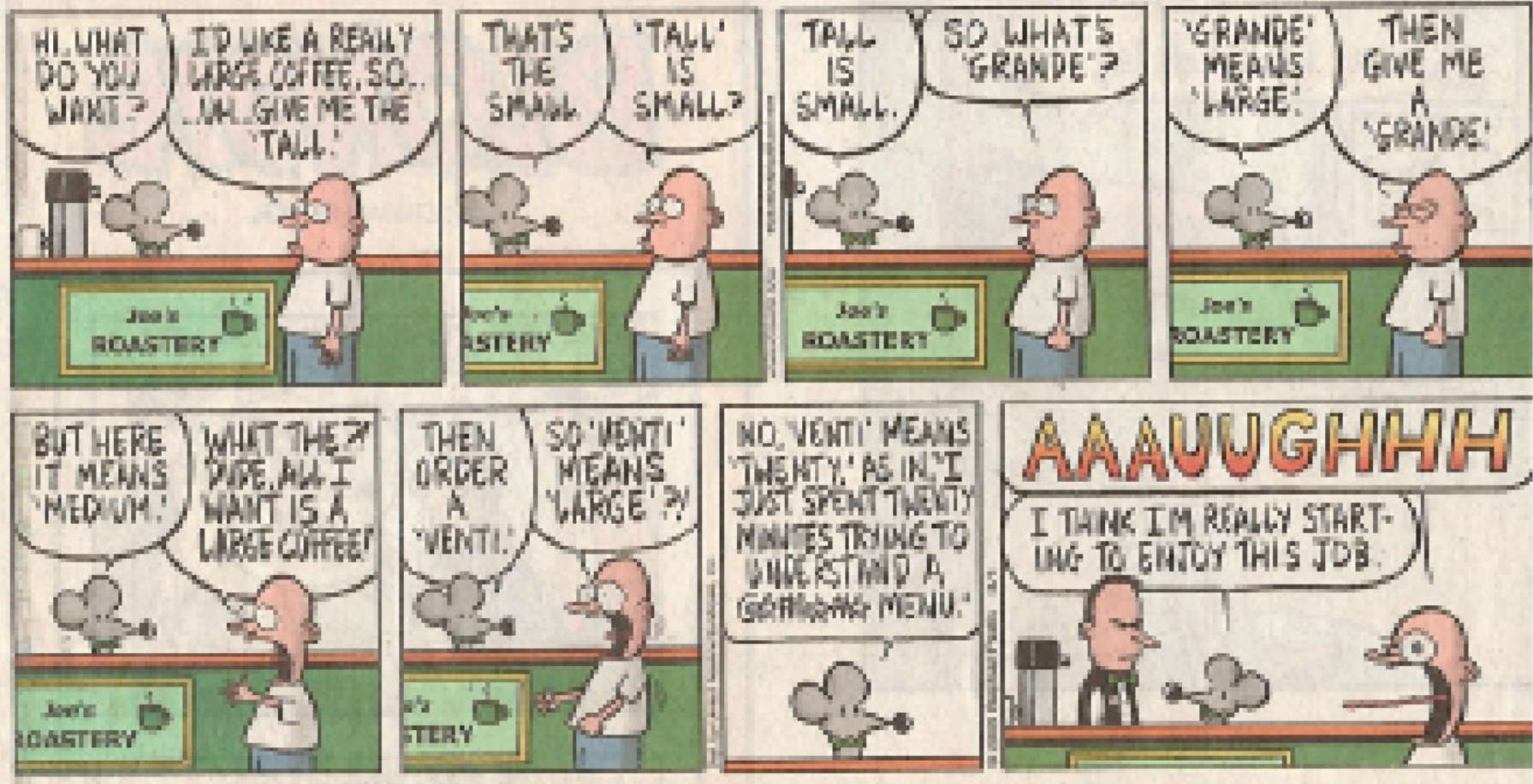
**RESOURCE IDENTIFIER:** (ISBN) 0872879674



# VOCABULÁRIOS

# ○ Problema do Vocabulário

## PEARLS BEFORE SWINE *Stephan Pastis*





# ○ Problema do Vocabulário

- Pessoas usam um grande variedade de palavras para a mesma coisa ou conceito
- A maioria das pessoas – em especial projetistas de sistemas – se surpreendem com isso pois acham que suas escolhas são “naturais”
- Esta grande variabilidade na seleção de palavras é um fato inescapável que tem raízes na natureza da linguagem e na categorização

# Furnas: o problema do vocabulário

- Pessoas usam palavras diferentes para descrever as mesmas coisas
  - “If one person assigns the name of an item, other untutored people will fail to access it on 80 to 90 percent of their attempts.”
  - “Simply stated, the data tell us there is no one good access term for most objects.”

# ○ Problema do Vocabulário

- Como conseguimos nos entender?
  - Contexto compartilhado
  - Diálogo
- Como as máquinas podem entender o que dizemos?
  - Contexto compartilhado?
  - Dialogo?

# Como conseguimos nos entender?

- Pode parecer incrível que as pessoas consigam se comunicar visto que organizam e nomeiam o mundo de formas tão diferentes
- Estabelecemos um contexto compartilhado através do diálogo
- Utilizamos termos raros quando queremos ser precisos
- Os problemas básicos estão na forma de conceber e categorizar coisas e conceitos

# Vocabulários controlados

- Controle de vocabulário é uma tentativa de fornecer um conjunto consistente e padronizado de termos (ex: cabeçalhos, nomes, classificações, etc.) com a intenção de auxiliar o usuário a encontrar informação
  - Um conjunto consistente de descrições para serem usados como Metadados

# Vocabulário controlado

- Conjunto de termos padrão para determinado domínio ou contexto
- Exemplos de uso
  - Padronização de termos em BDs para, por exemplo, domínios de palavras chave
  - Descritores em ambientes de GED
  - Base para esquemas de classificação
  - Hashtags

# Controle de Nomes

- Cutter (1876): objetivos da descrição bibliográfica
  - Permitir a uma pessoa encontrar um documento do qual
    - O autor, ou
    - O título, ou
    - O assunto é conhecido
  - Mostrar o que uma biblioteca tem
    - De um dado autor
    - Sobre um dado assunto (e correlatos)
    - De um dado tipo (ou forma) de literatura.
- Primeiro serve ao acesso
- Segundo serve a co-localização

# Problemas com nomes

- Quantos nomes devem estar associados a um documento?
- Qual destes deve ser a “entrada principal?”
- Que forma cada nome deve ter?
- Que referencias devem ser feitas de outras formas de nome possíveis que não foram usadas?



# ○ problema

- Proliferação de formas de nomes
  - Nomes diferentes para a mesma pessoa
  - Pessoas diferentes com o mesmo nome
- Exemplo
  - Books in Print (semi-controlado)

# Goethe

## Books in Print

GOETHE, JOHANN

G's

- Goethe, J. W. Von see Von Goethe, J. W.
- Goethe, J. W. Von see Von Goethe, J. W. & Steiner, Rudolf.
- Goethe, Johann W. Von see Goethe, Johann Wolfgang Von.
- Goethe, Johann W. Von see Goethe, Johann Wolfgang von.
- Goethe, Johann W. Von see Goethe, Johann Wolfgang Von.
- Goethe, Johann W. von see Von Goethe, Johann W.
- Goethe, Johann Wolfgang Von. The Autobiography of Johann Wolfgang von Goethe. Vol. I. pap. 15.00 (ISBN 0-226-30057-9, Phoen); Vol. II. pap. 15.00 (ISBN 0-226-30058-7, P603). U of Chicago Pr.
- The Autobiography of Johann Wolfgang Von Goethe. Oxford, John, tr. from Ger. 1975. Vol. II. 15.00 (ISBN 0-226-30056-0). U of Chicago Pr.
- Autobiography: Truth & Fiction Relating to My Life, 10 vols. Oxford, John, tr. 1985. Repr. of 1901 ed. Set. lib. bdg. 500.00 (ISBN 0-8492-2836-0). R West.
- Dr. Faust: An Art Book, 2 vols. (Illus.). 237p. 1984. Repr. of 1908 ed. Set. 217.55x (ISBN 0-89901-161-6). Found Class Reprints.
- Dr. Henry Faust: The Tragedy's First & Second Parts. Roth, John A., tr. from Ger. LC 87-18910. 284p. (Orig.). 1987. lib. bdg. 26.25 (ISBN 0-8191-6554-9); pap. text ed. 14.25 (ISBN 0-8191-6555-7). U Pr of Amer.
- Egmont. Trask, Willard, tr. from Ger. LC 60-14229. (Orig.). 1960. pap. text ed. 5.95 (ISBN 0-8120-0060-9). Barron.
- Egmont. Waidson, H. M., ed. (German Text Ser.). 146p. 1974. pap. 15.95x (ISBN 0-631-01670-8). Basil Blackwell.
- Egmont: A Play. Passage, Charles E., tr. from Ger. 110p. 1985. pap. 5.95x (ISBN 0-8044-6185-6).

- Goethe, the Lyricist. 2nd. rev. ed. LC 78-31302. (BCL Ser.: No. 1). Repr. of 1955 ed. 27.00 (ISBN 0-404-50916-9). AMS Pr.
- Goethe's Autobiography. 1949. 15.00 (ISBN 0-8183-0224-0). Pub Aff Pr.
- Goethe's "Faust". LC 61-5971. (Anchor Literary Library). 1963. pap. 7.95 (ISBN 0-385-03114-9, Anch). Doubleday.
- Goethe's "Faust". Fairley, Barker, tr. LC 74-151823. (Illus.). 1970. pap. 14.95 (ISBN 0-8020-6153-2). U of Toronto Pr.
- Goethe's Plays. Passage, Charles E., tr. from Ger. LC 79-4834. 1980. 30.00x (ISBN 0-8044-2258-3). Ungar.
- Goethe's Theory of Colours. lib. bdg. 79.95 (ISBN 0-87968-196-9). Gordon Pr.
- Goethe's World View: Presented in His Reflections & Maxims. Ungar, Frederick, ed. Norden, Heinz, tr. LC 63-18513. pap. 5.95x (ISBN 0-8044-6192-9). Ungar.
- Gotz Von Berlichingen. Passage, Charles E., tr. LC 65-16623. (Orig.). pap. 7.95x (ISBN 0-8044-6187-2). Ungar.
- Hermann & Dorothea. Coogan, Daniel, tr. LC 66-21737. (Bilingual). pap. 7.95x (ISBN 0-8044-6188-0). Ungar.
- Iphigenia in Tauris. Passage, Charles E., tr. pap. 7.95x (ISBN 0-8044-6189-9). Ungar.
- Italian Journey: 1786-1788. Auden, W. H. & Mayer, Elizabeth, trs. from Ger. LC 81-86248. 528p. 1982. pap. 16.50 (ISBN 0-86547-076-6). N Point Pr.
- Johann Wolfgang von Goethe's Roman Elegies & Venetian Epigrams. Lind, L. R., tr. from Ger. LC 74-7172. xii, 228p. 1974. 25.00x (ISBN 0-7006-0125-2). U Pr of KS.
- Kindred by Choice. Waidson, H. M., tr. from Ger.

V's

- Ballinger Pub.
- Von Gloeden, Wilhelm, photos by. Taormina. (Illus.). 112p. 1986. 50.00 (ISBN 0-942642-22-8). Twelvetroes Pr.
- Von Guclinski, Stefan, ed. Liberia in Maps. LC 72-80411. (Graphic Perspectives of Developing Countries Ser.). (Illus.). 111p. 1973. 35.00 (ISBN 0-8419-0126-0, Africans). Holmes & Meier.
- Von Goethe, J. W. Conversations with Eckermann. Oxford, John, tr. from Ger. 384p. (Orig.). 1984. pap. 16.50 (ISBN 0-86547-148-7). N Point Pr.
- Von Goethe, J. W. & Steiner, Rudolf. The Fairy Tale of the Green Snake & the Beautiful Lily. 2nd ed. LC 78-73644. 72p. (Orig.). 1981. pap. 3.50 (ISBN 0-89345-203-3, Steinerbks). Garber Comm.
- Von Goethe, J. W. see Goethe, Johann Wolfgang Von.
- Von Goethe, Johann see Goethe, Johann Wolfgang Von.
- Von Goethe, Johann W. Goethe, Johann Wolfgang von, Italian Journey. Saine, Thomas P. & Sammons, Henry, eds. Heitner, Robert P., tr. from Ger. (Goethe's Collected Works: Vol. 6). 1988. 32.50 (ISBN 0-317-67565-6). Suhrkamp.
- The Sorrows of Young Werther & Novella. Mayer, Elizabeth, et al, trs. LC 84-4592. 201p. 1984. 7.95 (ISBN 0-394-60509-8). Modern Lib.

...etc...

# Usos de vocabulários controlados

- Entradas de assunto de biblioteca e classificação
- Indexação comercial de revistas e bases de dados
- Yahoo! e outros esquemas de classificação na web
- Sistemas online e manuais dentro de organizações





# CATEGORIZAÇÃO

# O que são categorias?

- Categorias são *classes de equivalência* – conjuntos de coisas materiais e abstratas, processos e eventos que tratamos da mesma forma
  - Isso não significa que toda instância de uma categoria seja exatamente igual, mas que, sob alguma perspectiva ou objetivo, podemos tratá-las como equivalentes, com base em algo que têm em comum
- Categorias são modelos cognitivos (e lingüísticos) que permitem a aplicação conhecimento prévio a novas situações e reconhecer instâncias ou eventos
- Usamos categorias sempre que nos comunicamos analisamos, prevemos ou classificamos

# Tipos de sistemas de categorização

- Categorização Cultural
- Categorização Individual (“Tagging”)
- Categorização Institucional (“Semântica de Negócio”)

# Princípios da Categorização

- Similaridade
- Hierarquia
- Alguns níveis em uma hierarquia de categorias parecem mais “naturais” ou “básicos” que outros
  - Mas categorias podem ser não hierárquicas e ter sobreposição
  - Mas podemos criar categorias completamente novas prontamente
- A maioria das coisas pode ser categorizada em um número grande de formas e qualquer característica de um objeto pode ser usada para fazê-lo



# Por que estudar categorização?

- Categorização é central à forma como organizamos informação (e o mundo)
- É um processo cognitivo importante
- Entender como as pessoas categorizam as coisas pode nos auxiliar a projetar sistemas melhores ao organizar e recuperar informação
  - Recentemente, categorizações centenárias foram revistas



# TAXONOMIA

# Taxonomia: conceito

- As taxonomias são ferramentas de organização intelectual, atuando como um mapa conceitual de um domínio de atividades, possibilitando a recuperação de dados e informações através de navegação

# Taxonomias

- Definição: “vocabulário controlado e dinâmico de descritores relacionados com semântica e genericamente, que cobre de forma extensiva um ramo de conhecimento”

*(Dicionário Aurélio)*

- O termo taxonomia vem do grego:  
“taxis”, que significa “arrumação” e  
“nomos”, que significa “lei”

# Taxonomia

- Classificação de elementos em um domínio
  - Classificação
    - Operação de agrupamento de elementos e estabelecimento de relações entre eles, de forma sistêmica
  - Domínio
    - Uma esfera de conhecimento ou atividade

# Taxonomia

- Conceito
  - Esquema classificatório, uma lista estruturada de categorias de assunto, como um Mapa Conceitual. Apresenta as relações hierárquicas entre os tópicos.
- Função da Taxonomia
  - Possibilita a organização e acesso à informação através de NAVEGAÇÃO.

# Exploração e Produção de Petróleo

## **Exploração**

**Estudos de Reservatórios**

**Avaliação das Formações**

**Estimativa de Reservas**

**Construção/Engenharia de Poços**

**Perfuração**

**Completação e Restauração**

**Controle de Poço**

**Cimentação Secundária**

**Operações com Fluidos**

**Operações Especiais**

**Limpeza de Poço**

**Abandono de Poço**

**Desenvolvimento da Produção**

**Produção**

**Elevação e Escoamento**

**Coleta e Separação de Óleo, Gás e Água**

**Tratamento de Fluidos Produzidos**

**Processamento de Óleo e Gás**

**Estocagem, Movimentação e Transferência**

Processos

# Taxonomias na Gestão do Conhecimento

- Categorias organizadas que permitem a localização de informações e dão suporte à descoberta de conhecimento
- Suporte à acesso via navegação
  - Complementa mecanismos de busca
    - Na busca, preciso saber o que procuro
  - Na navegação via taxonomia, é possível ver o que está disponível



# Taxonomia - Relacionamentos

- Classifica os termos hierarquicamente
- Relacionamentos pai-filho apenas
  - Generalização (IS-A)
  - Part-of, causa-efeito, associação... não são usualmente permitidos

# Taxonomia e Base Terminológica



Uma Taxonomia pode conter um vocabulário controlado associado

**Taxonomias com vocabulários associados  
são consideradas Bases terminológicas**

# Taxonomias: aplicação

- O desenvolvimento de taxonomias para empresas tem sido um dos pilares da gestão da informação e do conhecimento
  - O volume de informação cresce a cada dia, para viabilizar um acesso rápido e eficiente às informações é necessária a utilização de meios de organização e recuperação de informações padronizadas

# Taxonomias: aplicação

- O uso de taxonomias permite que se estabeleça padrões de alto nível para a ordenação e classificação de informação, além de permitir que as organizações reconheçam e relacionem atividades, diminuindo esforços na produção e utilização do conhecimento

# Captura do Conhecimento

- A captura do conhecimento é uma etapa da elaboração da taxonomia em que o objetivo é o levantamento dos termos que deverão ser acomodados numa estrutura classificatória, servindo como ponto de acesso à informação
- O levantamento dos termos/assuntos pode-se dar através de:
  - análise dos documentos existentes na instituição;
  - entrevistas com especialistas da área;
  - outros instrumentos classificatórios ou terminológicos.

# Concluindo...

- Taxonomias são Ferramentas de Gestão e Recuperação de Informações
- Reflete as “inteligências”(potenciais de conhecimento) de uma organização
- Reflete o campo de atuação de uma organização, possibilitando recuperação de dados e informações através de navegação.

# Concluindo ...



- A organização das informações através do conceito de Taxonomia permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica

