

Sistemas de Organização do Conhecimento



PROFA. LILLIAN ALVARES

FACULDADE DE CIÊNCIA DA INFORMAÇÃO

UNIVERSIDADE DE BRASÍLIA



- Há concordância dos autores quanto à **importância dos SOC para a organização e recuperação de informações...**
 - ✦ Mas, parece não haver tanto consenso quanto à **abrangência do conceito** de SOC e aos tipos de sistemas considerados.



- Vickery ressalta que, com o tempo, esses sistemas **tornaram-se mais complexos** e assumiram **funções mais amplas**, havendo, em consequência, o **surgimento de novas denominações**, como:

Tipos



Pré-coordenadas

- Sistemas de classificação
- Listas de cabeçalhos
- Glossários
- Dicionários
- Entre outros

Pós-coordenadas

- Tesouro
- Taxonomias
- Ontologias
- Redes Semânticas
- Entre outros

Autores



Salton (1990)



- Afirma que qualquer **Sistema de Organização do Conhecimento...**
 - ✦ ... deve lidar com o problema da **análise da linguagem...**
 - ... uma vez que o **conteúdo dos textos e dos documentos necessariamente controla as atividades de recuperação.**

Broughton et al, 2004



- SOC são ferramentas semânticas que consistem de
 - ✦ **conceitos e**
 - ✦ **relações semânticas,**
 - definidas e selecionadas.

SKOS (Simple Knowledge Organization System), 2004



- É um conjunto de elementos, geralmente **estruturado e controlado...**
 - ✦ ... que pode ser usado para **descrever objetos**, navegar em coleções, etc.

Hjorland, 2008



- Ferramentas que apresentam a **interpretação organizada de estruturas do conhecimento**, também chamadas de **ferramentas semânticas**.
 - ✦ Contemplam, essencialmente, **conceitos e suas relações semânticas**.

Lima, 2007



- **Organização do conhecimento pela elaboração da **informação documentária**,**
 - ✦ Isso é, pelas **linguagens documentárias**

Informação e Linguagens Documentárias



Contexto



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

Contexto



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



- A informação documentária pode ser considerada, portanto, o **elo entre o documento e o usuário** e...
 - ✦ ... como toda representação
 - ... não é o próprio objeto (documento).



- Ao entendermos que toda informação produzida visa à sua efetiva utilização...
 - ✦ ... temos na atividade de organização da informação um **papel fundamental de caráter mediador...**
 - já que essa atividade é responsável **pela comunicação entre a produção e o uso de informações.**



- Trata-se do resultado de um processo de **tradução semântica daquilo que é dito pelo autor...**
 - ... por uma **representação do conteúdo do documento...**
 - ✦ ... levando em consideração fatores condicionantes como **necessidade do usuário, domínio tratado,** entre outros.

Contexto



- É uma tradução que permite a **transformação de informações dos textos para uma linguagem documentária controlada,**
- Envolve a questão da significação:
 - **Como representar adequadamente as informações sem comprometer o seu significado?**

Características da tradução



Contexto



- 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.

Definições



Definições



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

Definições



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

Definições



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

Definições



- Para a escolha do termo correto para o descritor a fim de **reduzir a diversidade e a ambiguidade da terminologia** e estabelecer 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.**

Terminologia



- Linguagem Documentária, também conhecida como:
 - **Linguagem Controlada**
 - **Linguagem de Indexação**
 - **Linguagem Descritoras**
 - **Codificações Documentárias**

Linguagens Pré-Coordenadas e Pós-Coordenadas



Linguagens Pré-Coordenadas



Linguagens pré-coordenadas



- **Combinação feita na etapa de entrada das palavras**
- 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



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

Linguagens pré-coordenadas



- Não são flexíveis, pois não permitem busca por qualquer termo.

Linguagens pré-coordenadas



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

Desvantagens



- Necessitam de múltiplas entradas para explicitar todos os conceitos
- Difíceis de atualizar, pouco dinâmicas
- Menor flexibilidade na indexação
- Relações estáticas
- Menor revocação, menor precisão

Exemplos



- Índices
- Listas de cabeçalhos de assuntos
- Glossários
- Dicionários

Linguagens Pós-Coordenadas



Linguagens pós-coordenadas



- Combinação feita no **momento da saída**, na **recuperação da informação**.
- Utilização do conceito 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
- Maior flexibilidade na indexação

Desvantagens



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

Exemplos



- Taxonomias
- Ontologias

Classificação de SOC



Por Vickery, 2008



Característica:

SOC cumprem diferentes funções, de acordo com o **momento da evolução dos sistemas de informação.**

Portanto, é cronológica.

Por Vickery, 2008



1) Era da pré-coordenação:

Eram estruturas **estáticas** e atendiam às necessidades dos **sistemas manuais** de organização e recuperação da informação

Exemplos: **índices e catálogos.**

Por Vickery, 2008



2) Era da pós-coordenação:

Mais **dinâmicos** e possibilitam que **cada um de seus elementos** (termos) sejam manipulados de forma **independente** para representar os assuntos de cada documento.

Exemplos: vocabulários controlados e **tesauros**.

Por Vickery, 2008



3) Era da Internet:

Classificações hierárquicas que orientam o usuário na **escolha do termo** que melhor expressa sua questão de busca

Exemplo: **taxonomias**

Por Vickery, 2008



4) Era da Web Semântica:

São projetados por **agentes inteligentes**.

Exemplo: **ontologias**

Por Hodge, 2000



1) Listas de termos (geralmente acompanhados de suas definições)

- Lista de autoridades: controlam as variações de nomes para entidades, como nomes de países, indivíduos ou instituições.
- Glossários: lista de termos com definições, geralmente de um assunto ou domínio específico.
- Dicionários: lista alfabética de palavras e suas definições.
- Gazetteers: dicionário de nomes de lugares e acidentes geográficos, tais como cidades, rios, vulcões.

Por Hodge, 2000



2) Classificações e categorização: enfatizam a criação de classes de assuntos.

- Cabeçalhos de assunto: conjunto de termos controlados que representam os assuntos de uma coleção
- Classificação, taxonomias e esquemas de categorização

Por Hodge, 2000

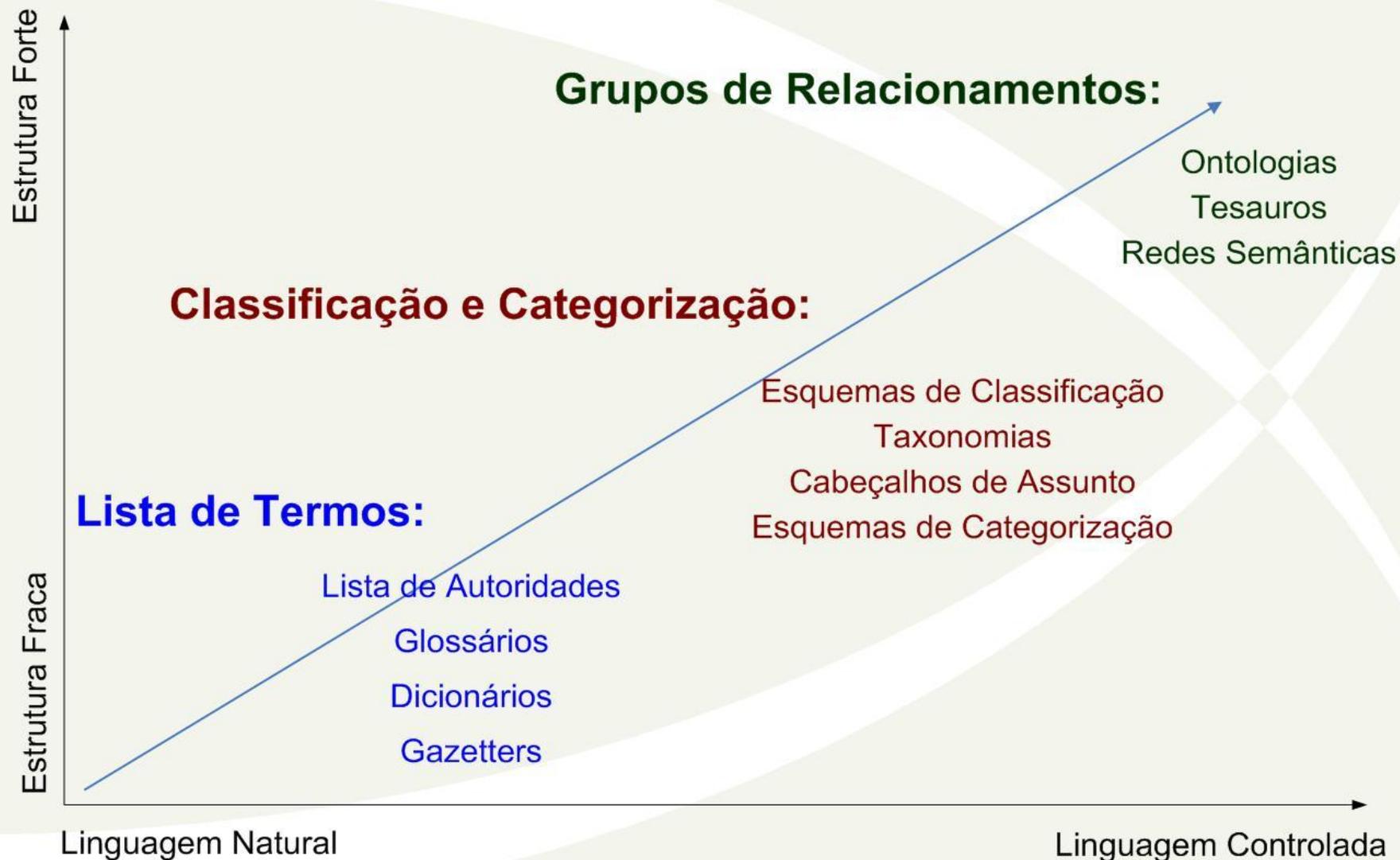


3) Grupos de relacionamentos: conexão entre termos e conceitos.

- Tesouros: conjunto de termos que representam conceitos e as relações de equivalência, hierárquicas e associativas que se estabelecem entre eles.
- Redes semânticas: estrutura de conceitos e termos em forma de rede ou teia, os conceitos são nós e os relacionamentos expandem-se a partir dos nós.
- Ontologia: estrutura de conceitos e representação dos relacionamentos complexos entre eles

Tipos de SOC por Hodge

(Fonte: Zeng & Salaba: FRBR Workshop, OCLC 2005)



Funções do SOC



Soergel (1999)



Auxílio à recuperação da informação

Fornecem uma base de conhecimento que apoia a **busca e recuperação de informação pelo usuário** final, como, por exemplo:

- menus em árvores
- busca por facetar
- cadeias hierárquicas
- mapeamento de termos de busca

Vickery (2008)



Explicitam o conhecimento de um domínio e possibilitam:

- organização e recuperação de informações ou de outros tipos de objetos
- construção de mapas de conhecimento
- navegação para busca de informação
- criação de novo conhecimento a partir do existente

Vickery (2008)



Auxílio à busca

Buscas genéricas:

Reúne todos os assuntos específicos em um campo de assunto mais geral, de maneira que possam ser encontrados pelo usuário

Buscas específicas:

Esses assuntos normalmente são descritos por uma combinação de termos, os quais devem estar reunidos de forma compreensível em um SOC

Vickery (2008)



Auxílio à busca

Arranjo sequencial:

arranjo de um conjunto de itens armazenados ou selecionados de maneira compreensível para o usuário

Escolha de termos de busca:

apresentação de termos semanticamente relacionados, de maneira que o usuário possa selecionar aqueles que melhor representem o assunto que deseja.

Shiri e Molberg (2005)



Prover várias perspectivas de uma coleção no ambiente digital

- Pois oferece ao usuário uma estrutura conceitual rica e uma estrutura semântica que facilitam a:
 - seleção de termos para a formulação, reformulação e expansão da busca.

Hodge (2000)



Em bibliotecas digitais

- Estabelecer elos entre recursos digitais e entre esses e objetos físicos.
- Os registros podem ser localizados diretamente por meio da navegação nos SOC ou indiretamente, por meio de uma chave de busca que pode ser usada para acessar outra fonte de informação.

McCulloch e Macgregor (2008)



Mapeamento terminológico entre diferentes SOC em repositórios digitais

- Pois com o crescimento do número de repositórios, objetos digitais são indexados e organizados segundo uma variedade de diferentes esquemas.
- É necessário, portanto, que os usuários possam consultar e navegar simultaneamente múltiplos repositórios distribuídos.

Referências



- 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.
- BRÄSCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. In: Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento*. Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capítulo 8, p. 147-176. Edição eletrônica. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).