

Debora Tiemi
2015

Indexação automática e recuperação da informação

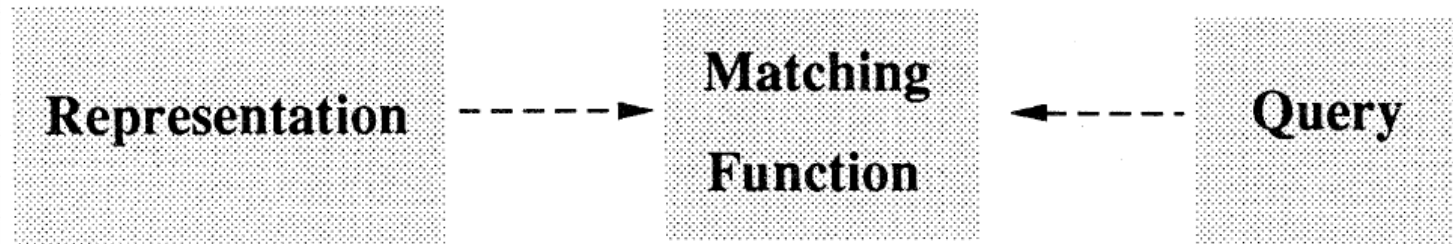
Proposta do seminário

- 0 Importância e relação da indexação e recuperação de informação
- 0 Mostrar o que é e como funciona a indexação automatizada para textos
- 0 Vantagens e desvantagens

Indexação

- 0 Análise documentária para representação temática
- 0 Segundo Lancaster (2004) a indexação possui duas etapas:
 - 0 Análise conceitual
 - 0 Tradução

Relação de indexação com recuperação da informação



INGWERSEN (1992)

- 0 Temacidade
- 0 Representação (tipos)
- 0 Relevancia e avaliação

Indexação automatizada

- 0 O que é

- 0 Como funciona

- 0 Indexação por extração automática

- 0 Indexação por atribuição automática

Indexação automatizada

- 0 Indexação assistida
- 0 Indexação semi-automática
- 0 Indexação automática



PRECIS

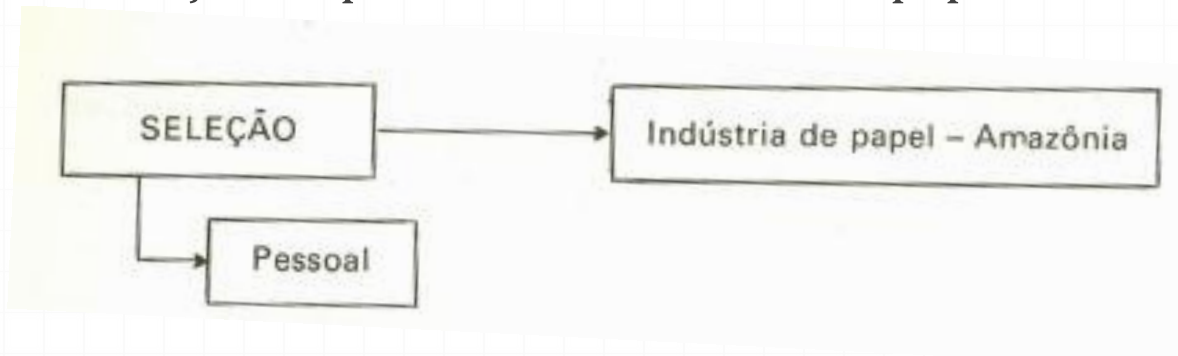
- 0 Preserverd Context Index System - PRECIS
- 0 British National Bibliography (BNB)
- 0 1974
- 0 Bibliotecário Derek Austin
- 0 Índice de assuntos

- 0 Indexação por atribuição automática
- 0 Indexação assistida

- 0 FUJITA, M.S.L. Sistema de indexação PRECIS. **Revista brasileira de Biblioteconomia e Documentação**, São Paulo, v.21, n.3/4, jul./dez.1988

PRECIS

- 0 Construção de cadeia de termos. Feito em 3 etapas:
- 0 Enunciado de assunto é fragmentado em conceitos separados.
- 0 Ex. Seleção de pessoal em indústria de papel na Amazônia



PRECIS

- 0 Determinação de operadores que expressam as funções desempenhadas pelos termos que representam conceitos (Realização manual pelo indexador)

Operadores principais
ambiente de sistema
observado

Sistema observado (núcleos
de operadores)

O. Localização

1. Sistema-chave: objeto de uma ação transitiva; agente de uma ação intransitiva;
2. Ação/efeito
3. Agente de uma ação transitiva: aspectos; fatóres.

A

Dados relatados p/observador

Exemplos selecionados

Apresentação de dados

Operadores interpostos
elementos dependentes

Conceitos de ligação

Conceitos coordenados

B

Operadores de diferenças
termos compostos
(prefixados por \$)

Conectivos
(componentes de ligação das
frases prefixadas por \$)

C

Ligação de temas
Ligação entre os termos, elos

4. Ponto de vista como forma;

5. Amostra de população/região de estudo;

6. Objetivo/forma;

p. Parte/propriedade
q. Membro de um grupo quase-genético (classe não-hierárquica);
r. Agregado

s. Definidor de função;
t. Associação atribuída pelo autor;

g. Conceito coordenado;

h. Diferença direta não-guia;
i. Diferença direta-guia;
j. Diferença a ser evidenciada;
k. Diferença indireta não-guia;
m. Diferença indireta guia;
n. Diferença não-guia a ser explicada;
o. Diferença guia a ser explicada;
d. Data, como uma diferença.

v. conector de leitura descendente.
w. Conector de leitura ascendente.

x. Primeiro elemento no tema coordenado;
y. Elemento subsequente no tema coordenado;
z. Elemento de tema comum.

PRECIS

0 Estabelecida organização dos termos em cadeia

AMAZÔNIA

Indústria de papel. Pessoal. Seleção

INDÚSTRIA DE PAPEL, Amazônia

Pessoal. Seleção

PESSOAL. Indústria de papel. Amazônia

Seleção

SELEÇÃO. Pessoal. Indústria de papel. Amazônia.

PRECIS

- 0 Artigos noticiosos
 - 0 Maior grau de dificuldade na determinação do assunto
 - 0 Cadeia de termos construída mais de uma vez
- 0 Artigos de vulgarização científica
 - 0 Vocabulário preciso
 - 0 Menor grau de dificuldade na construção da cadeia
- 0 Artigo de conteúdo especializado
 - 0 Enunciado de assunto se baseou no título
 - 0 Grau mínimo de dificuldade
 - 0 Cadeias maiores
 - 0 Maior número de entradas

SISA

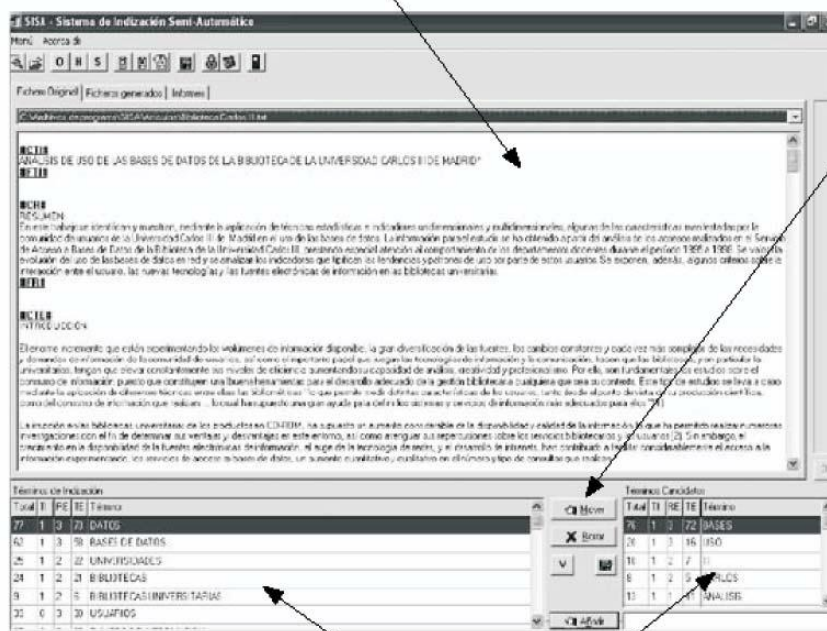
- 0 Sistema de Indización Semi-automático (SISA)
- 0 Universidade de Murcia (Espanha)
- 0 1999
- 0 Prof. Isidoro Gil Leiva
- 0 Indexação por extração automática
- 0 Indexação semi-automática
- 0 NARUKAWA, C.M. GIL LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de periódicos científicos. **Informação & Sociedade**: Estudos, João Pessoa, v.19, n.2, maio/ago. 2009.

SISA

- 0 O processo de indexação acontece em 3 módulos:
 - 0 Módulo 1: Sinalização do documento (título, resumo e texto) com marcadores
 - 0 Módulo 2: Análise do conteúdo. Busca e seleção de termos.
 - 0 Módulo 3: Aplicação de um critério de avaliação para seleção final dos termos.

SISA

Espacio destinado al texto objeto de la indización y también para la edición y/o validación del resultado de SISA, si fuera preciso.



Los iconos Mover, Borrar y Añadir sirven para la edición de la indización propuesta por SISA. En cambio, los iconos representados por una v y un disquete sirven, uno, para navegar por el vocabulario controlado y, otro, para exportar la indización a la base de datos.

Se muestran los términos de indización y los términos candidatos obtenidos por SISA, así como sus frecuencias y lugar de aparición (título, resumen o texto).

SISA

- 0 Análise dos resultados:
- 0 Acentos
- 0 Singular/plural
- 0 Termos simples / termos compostos,
- 0 Pequenas variações de termos compostos da/de
- 0 Termo autorizado apenas em uma parte do texto
- 0 Termos relacionados com termo relevante

BR/Search

- 0 Mineração de texto
- 0 Indexação por atribuição e extração automática
- 0 Indexação automática
- 0 ARAUJO, JÚNIOR, R.H.de. **Precisão no processo de busca e recuperação da informação.**
Brasília : Thesaurus, 2007

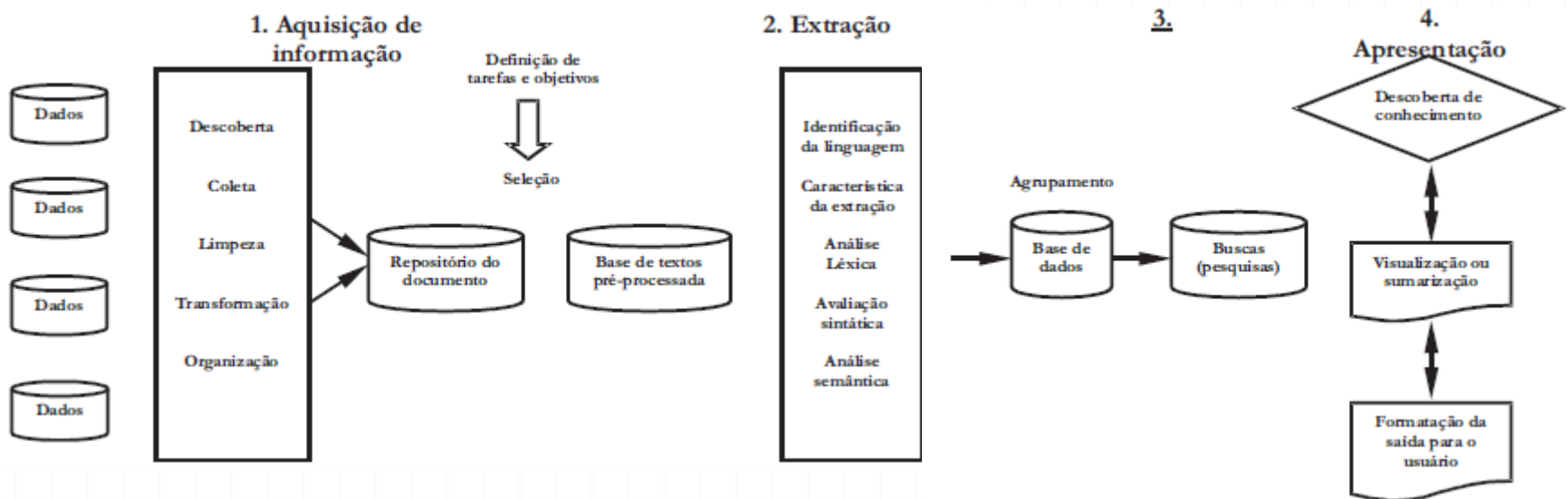
BR/Search

0 Mineração de textos:

- 0 Extração de informações sobre tendências e padrões em grandes quantidades de texto contidas em bases textuais.
- 0 Textos em linguagem natural.
- 0 Classificação automática de textos
- 0 Agrupamento por semelhança

BR/Search

0 Etapas do processo



BR/Search

- 0 Como foi feita a pesquisa
- 0 56 teses e dissertações em .pdf na base de dados Infohab

$$\text{Precisão} = \frac{\text{Número de documentos úteis recuperados pelo sistema}}{\text{Número total de documentos encontrados pelo sistema}} \times 100$$

- 0 Usuários (especialistas) do sistema avaliando se itens recuperados foram considerados útil ou inútil.

BR/Search

- 0 Indexação com mineração de texto x indexação manual
 - 0 Não houve ganho significativo no processo de busca e recuperação da informação
 - 0 Quantidade de itens bibliográficos maior do que a lista de palavras-chave utilizadas na indexação manual
- 0 Instrumento para enriquecimento de lista de palavras-chave, construção de vocabulário controlado

Vantagens

- 0 Alguns fatores apresentados por Méndez Rodriguez e Moreiro González (1999):
 - 0 Alto custo da indexação humana
 - 0 Aumento de informação eletrônica e proliferação de texto completo
 - 0 Gestão eletrônica de documentos (GED) e informatização do processos documentais
 - 0 Automatização dos processos cognitivos e avanços em Processamento de Linguagem Natural

Desvantagens

- 0 Diferentes significados em diferentes contextos, relacionar e selecionar conceitos implícitos
- 0 Indexação automática reconhece palavras e não conceitos
- 0 Normalmente restringem-se às áreas específicas do conhecimento
- 0 Dificuldade de se conseguir indexação totalmente automática

Referências

- 0 ARAUJO, JÚNIOR, R.H.de. **Precisão no processo de busca e recuperação da informação.** Brasília : Thesaurus, 2007
- 0 FUJITA, M.S.L. Sistema de indexação PRECIS. **Revista brasileira de Biblioteconomia e Documentação**, São Paulo, v.21, n.3/4, jul./dez.1988. Disponível em: < <http://www.brapci.ufpr.br/documento.php?dd0=0000002144&dd1=f3345> >. Acesso em: 02 jun. 2015
- 0 INGWERSEN, *Peter*. **Information Retrieval Interaction** . Michigan: Taylor Graham, 1992
- 0 LANCASTER, F. W. **Indexação e resumos:** teoria e prática. Brasília: Briquet de Lemos, 2004.
- 0 MENDEZ RODRIGUEZ, E. M; MOREIRO GONZÉLEZ, J. A. Lenguaje natural e indización automatizada. **Ciencias de la Información**, v.30, n.3, set. 1999. Disponível em: < <http://eprints.rclis.org/12685/1/indizacion99.pdf> >. Acesso em: 02 jun. 2015
- 0 NARUKAWA, C.M. GIL LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de periódicos científicos. **Informação & Sociedade: Estudos**, João Pessoa, v.19, n.2, maio/ago. 2009. Disponível em: < <http://www.brapci.ufpr.br/documento.php?dd0=0000007566&dd1=1c1e1> >. Acesso em: 02 jun. 2015