

Extracción de palabras claves de la tabla de contenido e índices analíticos

Ana M. Martínez Tamayo

[Ponencia presentada por:
Prof. Cristina Ristuccia y
Prof. Edgardo Stubbs]

Buenos Aires, 2008

Propósito de la ponencia

1. Mostrar la técnica de
 - Barbara Settel & Pauline Cochrane
 - Para extracción de palabras claves
2. Comparar con catálogos locales

La técnica

- Extraer palabras claves de
 - Tabla de contenido
 - Índice analítico
 - De libros y monografías
- Incorporar al registro bibliográfico
 - Términos no controlados o
 - Resumen indicativo o
 - Conceptos traducidos a descriptores

Definir la fuente de términos

Usar sólo la tabla de contenido

Si no es suficiente

Usar también el índice analítico

Si no sirve la
tabla de contenido

Usar sólo el índice analítico

Tabla de contenido

- ¿Tomo, parte, capítulo, apartado?
- Extraer sólo palabras significativas
- Eliminar: *prólogo, prefacio, anexo, etc.*
- (compilar un *stopword*)
- No repetir los términos seleccionados
- No modificar los términos
- (algunas excepciones)

Ejemplo

Prefacio	vii
Lista de figuras	viii
Parte 1. Teoría y descripción	
Capítulo 1. Introducción	1
Capítulo 2. Principios de <u>indización</u>	5
Capítulo 3. Práctica de la indización	20
Capítulo 4. <u>Indices pre-coordinados</u>	42
Capítulo 5. <u>Coherencia de la indización</u>	61
Capítulo 6. <u>Calidad de la indización</u>	75
Capítulo 7. <u>Resúmenes</u> : tipos y funciones	servicios de indización
Capítulo 8. La <u>redacción del resumen</u>	servicios de resúmenes
Capítulo 9. Aspectos de la <u>evaluación</u>	
Capítulo 10. Métodos adoptados en <u>servicios de indización y resúmenes</u>	143
[...]	

Índice analítico de materia

- Seleccionar:
 - Términos con ≥ 5 páginas consecutivas
 - Términos con ≥ 10 páginas no consecutivas
 - Términos con ≥ 5 subdivisiones

Ejemplo

Diccionario literario de obras y personajes, 339, 341

Diccionarios, 85-129

alemanes, 120-122

bibliografía, 99

bilingües, 124-129

de americanismos, 106-108

de argentinismos, 108-111

españoles, 99-106

franceses, 111-115

historia, 85-88

[...]

Estados Unidos. Congress. Library, 34, 99, 160, 162, 167, 168, 184, 226, 227, 229, 250, 237, 239, 250, 251, 253, 259, 313.

Diccionarios bilingües

Diccionarios españoles

Diccionarios franceses

Traducciones

- Monografías en otros idiomas
- Traducir los términos
 - Traducción directa por el indizador
 - Otros afectan la rapidez de la técnica

Representación

- Una de tres posibilidades:
 - Términos no controlados
 - Resumen indicativo
 - Traducción a descriptores

Términos no controlados

Campo 653 de Marc21

- #0\$a actualización
- #0\$a compatibilidad
- #0\$a construcción de un tesoro
- #0\$a control del vocabulario
- #0\$a evaluación de los tesauros
- #0\$a homografía
- #0\$a identificadores
- #0\$a lenguaje natural
- #0\$a listas de control
- #0\$a [...]

Resumen indicativo

Campo 520 de Marc21

3#\$a Este manual trata sobre actualización, compatibilidad, construcción de un tesoro, control del vocabulario, coste-beneficio, directrices, estructura del vocabulario, evaluación de los tesoros, homografía, identificadores, lenguaje natural, listas de control, normas, notas de aplicación, presentación del tesoro, procesamientos automáticos, recuperación de información, relación asociativa, relación jerárquica [...]

Traducción a descriptores

Palabras claves	Descriptores*
actualización	-
compatibilidad	-
control del vocabulario	control del vocabulario
directrices	directrices
identificadores	identificadores
lenguaje natural	lenguajes naturales
listas de control	catálogos de autoridades
procesamientos automáticos	automatización de tesauros
vocabulario de entrada	relaciones de equivalencia

*Tesoro de Bibliotecología y Documentación

Descriptores 2

Campo 650 de Marc21

#7\$a automatización de tesauros\$2tbd

#7\$a catálogos de autoridades\$2tbd

#7\$a control del vocabulario\$2tbd

#7\$a directrices\$2tbd

#7\$a elaboración de tesauros\$2tbd

#7\$a identificadores\$2tbd

#7\$a lenguajes de recuperación\$2tbd

#7\$a lenguajes naturales\$2tbd

#7\$a lenguajes postcoordinados\$2tbd

#7\$a [...]

Comparación con catálogos locales

Metodología

- Se aplicó la técnica a cinco libros:
 - Lancaster (1995a) Indización y res [...]
 - Lancaster (1995b) El control del voc [...]
 - Sabor (1978) Manual de fuentes de [...]
 - Sabor (1984) Manual de bibliotecol [...]
 - Dobra (1997) La biblioteca popular [...]
- Se comparó con catálogos de siete bibliotecas: A, B, C, D, E, F y G

Resultados

Libros	Palabras claves de TC		Descriptores asignados	
	Demora (min)	n	Biblioteca	n
Lancaster 1995a	1½	14	B,D,F,G	2,3,4,1
Lancaster 1995b	2	26	C,E,F,G	5,3,7,1
Sabor, 1978	2	15	A,B,D,G	1,2,4,1
Sabor, 1984	2½	33	B,F,G	5,4,1
Dobra, 1997	2½	23	G	1
Promedios	2,1	26,4		2,8

Lancaster (1995b): El control del vocabulario

- Técnica de Settel & Cochrane:
 - 26 palabras claves
 - 19 traducidas a descriptores (73%)
- Catálogos locales
 - 2,8 descriptores en promedio (1 a 7)
 - 10 descriptores entre todas (A/G)
 - 6 conceptos no están en TC o índice

Conceptos no específicos

- Bibliotecología
- Bibliotecología y ciencia de la información
- Control terminológico
- Descripción de documentos
- Informática
- Lingüística

Conclusiones

- La técnica de Settel & Cochrane:
 - Aumenta la exhaustividad
 - Aumenta la especificidad
 - Es rápida y versátil
 - Permite estandarizar los procedimientos