



**VNIVERSIDAD  
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

*Departamento de Biblioteconomía y Documentación*

**Resumen** para la presentación de la tesis doctoral

**Indización y control de terminología en bibliotecas  
de ensino superior politécnico en Portugal:  
El sistema del Instituto Politécnico de Portalegre**

Título original:

INDEXAÇÃO E CONTROLO DA TERMINOLOGIA EM BIBLIOTECAS  
DO ENSINO SUPERIOR POLITÉCNICO EM PORTUGAL:  
O SISTEMA NO INSTITUTO POLITÉCNICO DE PORTALEGRE

M<sup>a</sup> Catarina Barradas Martins

*Tesis dirigida por: Professora Doutora Carmen Caro-Castro*

---

Salamanca, 2013

## Sumario

Introducción.....	2
Objetivos.....	3
Metodología.....	5
Nivel 1: Instrumentos de apoyo à la indización en las Bibliotecas Politécnicas portuguesas (BESP) .....	5
Recogida de datos.....	5
Variables.....	6
Nivel 2: características de indización en las Bibliotecas del Instituto Politécnico de Portalegre (BIPP) .....	7
Recogida y depuración de datos .....	7
Selección de la muestra .....	8
Variables.....	9
Errores UNIMARC .....	10
Errores de sintaxis .....	11
Terminología .....	12
Encabezamiento específico.....	13
Fuentes de terminología .....	14
Análisis .....	14
Resultados.....	14
Nivel 1 – Instrumentos de apoyo à la indización en las BESP .....	14
Nivel 2 – Características de la indización en las BIPP .....	16
Errores UNIMARC .....	16
Errores de Sintaxis.....	18
Terminología .....	19
Encabezamiento Específico.....	22
Fuentes de terminología .....	22
Conclusiones.....	23
Propuesta de Modelo de Indización.....	26

## Introducción

En las bibliotecas académicas tradicionales, a pesar de querer acompañar los avances tecnológicos se continúa trabajando con base en los mismos preceptos y técnicas documentales que tienen marcado las últimas décadas. Los procesos de catalogación e indización de documentos exigen un buen conocimiento para que su aplicación sea correcta, y la conjugación de estas prácticas con los actuales ambientes de busca en la web, resulta en nuevas visiones y retos en el acceso a la información, y particularmente al nivel de la catalogación. Las bibliotecas ‘tradicionales’ deben adaptarse continuamente a esta nueva realidad, y las que encuentren mayores dificultades respecto a su modernización tecnológica, deberán por lo menos ofrecer servicios de calidad.

El actual contexto de los CRAI (Centro de Recursos para el Aprendizaje y la Investigación) como nuevo modelo de bibliotecas universitarias, hace con que el modelo tradicional sea renovado, implicando una nueva postura para el bibliotecario, que se debe comprometer con la construcción del catálogo online para una comunidad cada vez más exigente con la calidad de la información. El catálogo es una herramienta de gran importancia y debe satisfacer las necesidades del usuario. Como dice Villén-Rueda (2006), “se somos conscientes de la importancia del catálogo estamos en la dirección cierta”. La indización y recuperación por materias tienen una directa correspondencia, y las investigaciones tienen demostrado que políticas de indización inadecuadas se repercuten en la forma como los usuarios utilizan el catálogo. De los varios abordajes posibles en el área de la recuperación de información, en este trabajo se estudia la perspectiva del uso del lenguaje alfabético en la indización de materias, y la necesidad del control de los índices en el catálogo.

Los estudios y proyectos en esta área se desarrollan a un ritmo difícil de acompañar, por su complejidad y dispersión; este factor junto con la escasez de recursos humanos especializados, son los motivos que más pueden contribuir para dificultar el proceso de indización en las bibliotecas. Se trata de una tarea exigente y especializada para la cual muchas instituciones no están sensibilizadas.

El estudio de caso de nuestra investigación, que incluye cuatro bibliotecas en red, e están trabajando para ubicar sus catálogos bibliográficos para un proyecto de catálogo colectivo virtual, se hace una indización ad hoc porque las unidades de información evolucionaran de forma aislada; la falta de control de la terminología de los índices de materia en cada una de ellas, se revela un problema en el contexto de catálogo bibliográfico

El principal objetivo de este trabajo ha sido evidenciar la dificultad en el desarrollo del proceso de indización en bases bibliográficas. Los instrumentos normativos y de apoyo a esa tarea son dispersos e difíciles de utilizar para una gran parte de los profesionales. Los términos de indización suelen ser recogidos de variadas fuentes, lo que potencia enormemente la inconsistencia se no se hace el control de autoridades, poniendo así en causa una recuperación eficaz de información para los usuarios.

Nuestra investigación se divide en dos partes distintas: la primera identifica el uso de instrumentos normativos y de apoyo a la indización en las bibliotecas de enseñanza superior en Portugal; en la segunda se observan las características de los índices temáticos en el grupo de bibliotecas del Instituto Politécnico de Portalegre.

## **Objetivos**

Los puntos centrales de nuestra investigación son, en primer lugar, comprobar la práctica de la indización de acuerdo con la normativa adecuada en las bibliotecas politécnicas portuguesas, e en segundo lugar, verificar la calidad de los índices de materia en las bases bibliográficas de un grupo de bibliotecas, a través de la análisis de una muestra de terminología, de forma a comprobar que hay debilidades en el desarrollo del proceso de indización, y de que no hay un control sistematizado de la terminología. A través de estas hipótesis, se pretende demostrar que la utilización de instrumentos normativos es poco consistente, e que ese facto puede ter influencia en la creación de índices de calidad.

En nuestro estudio pretendemos, en primer lugar, identificar la utilización de instrumentos normativos en el proceso de indización, por parte de los profesionales en las bibliotecas politécnicas en Portugal. Nuestra hipótesis de partida respecto a este punto es que en la mayoría de estas bibliotecas se hace la tarea de indización en condiciones desfavorables, respecto a la utilización de instrumentos normativos e al control del lenguaje documental. Como el instrumento de referencia nacional es el sistema de indización SIPORbase, hay un especial interés en saber se sirve de apoyo à la construcción de la lenguaje controlada.

En una etapa siguiente se analiza la terminología de las bases bibliográficas de un grupo de bibliotecas para conocer las características de la indización, e así comprobar dificultades decurrentes de una indización no adecuada. Con el conocimiento de tales características, se pretende justificar una propuesta de mejora metodológica en el proceso de indización para ese grupo específico. Si en el primero nivel de investigación es posible saber sobre la utilización del SIPORbase, en este segundo nivel de las características, se verifica si la terminología está representada de acuerdo con el sistema.

Para que sea posible el conocimiento de esos factores, se establecieron los objetivos:

Ob1- conocer la situación de las bibliotecas politécnicas en Portugal, respecto a la utilización de instrumentos normativos e herramientas de apoyo à a indización;

Ob2 – identificar la situación de un grupo específico de bibliotecas (Instituto Politécnico de Portalegre) cuanto a las características de los índices de materia en los catálogos bibliográficos, para:

- a) conocer las características cuanto al uso del formato de intercambio UNIMARC;
- b) conocer las características cuanto al control de la terminología;
- c) conocer las características de la sintaxis;
- d) comprobar la consistencia de los índices (campo 6xx) con las normas establecidas en el SIPORbase;
- e) comprobar la coherencia de la terminología con otras fuentes de control.

En la parte del encuadramiento teórico, se estableció el objetivo de clarificar las principales cuestiones vinculadas al proceso de indización, haciendo un abordaje à la teoría de la indización e recuperación de información, y a la problemática de los catálogos bibliográficos online, y de su contexto de utilización en bibliotecas del ensino superior.

## **Metodologia**

La metodología para este trabajo se ha desarrollado de acuerdo con los dos niveles de observación, el primero a través de una encuesta en las instituciones politécnicas, el segundo a través de una análisis de los registros bibliográficos de las bibliotecas del Instituto Politécnico de Portalegre, en la que fueran definidos tres ejes centrales: el formato UNIMARC, la sintaxis de los encabezamientos de materia y la terminología.

### **Nivel 1: Instrumentos de apoyo à la indización en las Bibliotecas Politécnicas portuguesas (BESP)**

La red de ensino superior politécnico en Portugal es constituida por 98 Escuelas Superiores establecidas en 15 Institutos Politécnicos e algunas Universidades. Las unidades de información (bibliotecas, centros de documentación, etc.) que pertenecen a esas instituciones, se encuentran, en algunos casos, fusionadas en una biblioteca general, o en un sistema de catalogación centralizada (la razón de la diferencia entre el número de instituciones e el de unidades de información apuradas para la recogida de datos).

### **Recogida de datos**

En este nivel de observación, de carácter exploratorio, se realizó un levantamiento de las bibliotecas portuguesas, cuyos responsables se contactaran por e-mail con la

presentación del estudio e el pedido de colaboración. Se contactaran 56 instituciones de las cuales se apuró un total de 44 respuestas válidas, correspondiendo a 78,5% del universo inquirido.

Cuanto a la selección de muestra, en este nivel de observación fue posible seleccionar todo el universo de estudio, una vez que se contactaran todas las instituciones existentes.

### **Variables**

#### **I1 – Utilización de normativa en el proceso de indización**

*(Debe estar presente una norma como la NP 3715, la NP 4036<sup>1</sup>, o el sistema SIPORbase);*

*La referencia a instrumentos no considerados como tal, o respuesta que no permiten identificar un instrumento, son contabilizadas en la categoría ‘não especificado’).*

#### **I2 – Utilización de léxico(s) o tesoro(s) para apoyo à la indización**

*(La referencia a instrumentos no considerados como tal, o respuestas que no permitan identificar el instrumento, son contabilizadas en la categoría ‘não especificado’).*

#### **I3 – Utilización de un fichero de autoridades de materia ligado a la base bibliográfica**

*(Sí o No)*

### **Análisis**

Los datos de las respuestas de la encuesta sobre ‘utilización de instrumentos de apoyo à la indización’ se introducirán y trataran en el software SPSS *Statistics* 17.0, con lo cual se elaboraran las tablas de distribución de resultados.

---

<sup>1</sup> NP 3715:1989 Documentação. Método para análise de documentos, determinação do seu conteúdo e selecção de termos de indexação. NP 4036:1992 Documentação. Tesauros monolíngues: directivas para a sua construção e desenvolvimento.

## **Nivel 2: características de indización en las Bibliotecas del Instituto Politécnico de Portalegre (BIPP)**

El IPP dispone de cuatro unidades de información inseridas en sus Escuelas Superiores, e de un fondo documental<sup>2</sup> de acceso reservado. En nuestro estudio son observadas estas cinco bases bibliográficas, cuyos índices están disponibles a través de catálogos en línea.

### **Recogida y depuración de datos**

En este nivel se estudia la información en las bases bibliográficas, adelante designadas de *B1EDU*, *B2SAU*, *B3TGE*, *B4AGR*, *B5GER*, y que representan las principales áreas de conocimiento en el contexto de ensino superior, (Educación, Salud, Tecnología y Gestión, Agricultura e un área General). La recogida de datos se hace a partir de los registros UNIMARC bibliográfico, para el análisis de la terminología de los índices de materia en el bloco 6xx.

Se identificarán los campos de materia en los registros, lo que permitió reconocer, por un lado, los registros bibliográficos sin materia, y por otro lado, verificar los que tienen términos en el campo 606 (nombre común) e en otros campos del bloco 6xx (nombres propios).

De los 86813 registros en las bases, 44% no tienen materias, que representa un porcentaje mui elevado de documentos que no están accesibles en las búsquedas por materias. Los registros que contienen campos de materias (48644) representan unos 56%.

---

<sup>2</sup> Escola Superior de Educação, Escola Superior de Saúde, Escola Superior de Tecnologia e Gestão, Escola Superior Agrária de Elvas, Fundo documental Professor Fortunato Queirós.

Verificamos que 97,55 de los registros con materia pertenecen al campo UNIMARC 606 (nombres comunes), e los restantes, 2,5% pertenecen a otros campos del bloco 6xx, de (nombres propios).

El campo 606 concentra casi la totalidad de las materias. B1EDU, 99,9%, B2SAU 99,5%, B3TGE 92,5%, B4AGR 96,1%, B5GER 52,2%. Observando la distribución por bases, los campos del bloco 6xx representan, en algunos casos, valores casi nulos, respectivamente 0,0%, 0,4%, 7,4%, 3,8%, 47,7%. . Únicamente en la B5GER hay un equilibrio entre materias de nombres comunes y nombres propios.

### **Selección de la muestra**

Respecto a los registros del bloco 6xx, fue posible analizar la totalidad de los registros (1211), pelo que no fue necesario seleccionar una muestra. Al contrario, en los registros del campo 606 (nombre común) se analizó una muestra. La muestra estadística es un método usado con frecuencia para análisis de la calidad de bases de datos, e el tamaño de la muestra pude oscilar entre 200 e más de 2000 registros. El procedimiento mas objetivo para la selección de registros es la extracción al azar. No obstante, dependiendo de las características del estudio, es posible utilizar criterios de selección para criar muestras representativas, cómo, la fecha de creación de los registros, la materia de los documentos, las reglas o normas de descripción, etc. (Ortego e Bonal 2000).

Por razones operacionales, se optó por una técnica no aleatoria, la muestra intencional, formada por elementos seleccionados deliberadamente, “generalmente porque el investigador considera que tienen características típicas o representativas de la población” (Vicente *et al.* 1996). En nuestro caso seleccionamos un criterio cronológico de creación de registros. Se recogerán los registros de los años 2008, 2009 e 2010, pues consideramos que estos representan mejor las tendencias de la indización.

Así se apuró una muestra de 1573 registros, de los cuales resultaran 2218 encabezamientos de materia referentes al campo 606. De los 1211 registros recogidos del bloco 6xx resultaran 2218 encabezamientos de nombres propios (como materia). Fueran, así analizados un total de 3500 encabezamientos.

En relación al nivel de significación de la muestra, consultamos la tabla de Arkin y Colton, para determinación de una muestra de un universo finito, e verificamos que para un universo de 50 000 elementos, con un nivel de confianza de 99,7%, e con una margen de error de 4%, la tabla presenta el valor de 1368 elementos. Nuestra muestra tiene un valor de 1573 registros en un universo de 48644; así que consideramos su significación dentro de esos parámetros.

### **Variables**

La sistematización de variables para las características de la terminología de las cinco bases bibliográficas fue establecida a partir de evidencias identificadas de forma global en todas ellas. Las variables se definieron para tres ejes centrales: el formato UNIMARC, la sintaxis de los encabezamientos e la terminología.

En los nombres comunes se establecieron otras dos variables, una relacionada con el principio del encabezamiento específico (uno de los principios básicos consignado en SIPORbase); otra relacionada con la cobertura terminológica en las bases, puesto que la terminología suele ser recogida de diversas fuentes.

Las variables fueran definidas de acuerdo con la tipología de los encabezamientos, de nombres comunes y de nombres propios.

### **Encabezamientos de Nombres Comunes (UNIMARC 606)**

Respecto a la forma pueden ser de dos tipos:

- Non subdividido (NC1) cuando contienen un subcampo ^a  
(*ex.* Economía)

- Subdividido (NC2) cuando contiene una o más subdivisiones  
(*ex.* Economia -- [Relatórios] -- Portugal -- 2000)

Encabezamientos de Nombres Propios (UNIMARC 600, 601, 607 – persona, entidad, lugar<sup>3</sup>).

En el caso portugués, los encabezamientos de nombres propios obedecen a las “Regras Portuguesas de Catalogação”.

600 Persona	<i>Ex.:</i> Pessoa, Fernando, 1888-1935
601 Entidad	<i>Ex.:</i> Instituto Politécnico de Portalegre. Escola Superior de Saúde
607 Lugar	<i>Ex.:</i> Alentejo (Portugal)

### **Errores UNIMARC**

Se verifica se los campos y subcampos están inseridos de forma correcta.

(En los encabezamientos de nombres comunes)

#### **U1**

Términos que no corresponden a la etiqueta de campo del bloco 6xx

600:~aAnálise de dados      (*nombre común en el campo de nombre de persona*)

#### **U2**

Términos que no corresponden a las etiquetas de subcampos (^j ^x ^y ^z).

##### **U2^j**

606:~aCuidados veterinários^jAnimais de companhia      (*^j es un subcampo de forma*)

##### **U2^x**

606:~aDesenvolvimento regional^xAlentejo      (*el aspecto geográfico debe ser representado por ^y*)

---

<sup>3</sup> Es igualmente considerado el nombre de familia un nombre propio, que corresponde al campo 602 UNIMARC, pero no ocurrió en os registros.

(En los encabezamientos de nombres propios)

### U3

Términos que no corresponden a las etiquetas de subcampos (^a, ^f).

Para el ejemplo: ^aPessoa^bFernando^f1888-1935)

#### U3^a

600:^aFernando Pessoa *(el nombre debería ser representado por el apellido en ^a)*

#### U3^f

600:^aPessoa^bFernando *(la(s) fecha(s) del autor deberían ser representadas en ^f)*

### Errores de sintaxis

Se analiza la orden de campos y subcampos en la representación de los encabezamientos.

### S1

Se verifica en los encabezamientos de nombres comunes, si las subdivisiones están representadas conceptualmente por materia principal en ^a, e siguientes ^j, ^x, ^y, ^z, correspondiendo respectivamente a faceta de forma, materia, geográfica e cronológica.

606: ^aEstatísticas^xPoder de compra *(en este caso, el concepto principal es 'Poder de compra', e 'Estatísticas' un calificador de forma).*

### S2

Se verifica en los encabezamientos de nombres propios, en las subdivisiones de materia, que se siguen al encabezamiento de acuerdo con las RPC, cuya forma es analizada en la variable de terminología.

601:^aSociedade Portuguesa de Autores^z1985^xAniversários *(la subdivisión cronológica debería ser presentada al final)*

## Terminología

Se analiza la terminología de acuerdo con los índices de la BNP (Biblioteca Nacional de Portugal), que siguen las reglas del SIPORbase.

### T1

Términos que no coinciden con los índices, en cualquier campo o subcampo.

*Ejemplos:*

	<i>(no conforme)</i>	<i>(sí)</i>
T1^a	606: ^a Obesidade	^aObesidade (Animais)
T1^j	606: ^aDesign^jDicionário	^aDesign^jDicionários
T1^x	606: ^aEconomia^xEexercícios de economia	^aEconomia^jExercícios
T1^y	606: ^aSociologia rural^yAlentejo	^aSociologia rural^yAlentejo (Portugal)

### T2

Utilización de términos compuestos en los encabezamientos no subdivididos (Nombres Comunes)

En los encabezamientos ‘no subdivididos’, observamos la utilización de términos simples e compuestos; de acuerdo con las directrices SIPORbase estos encabezamientos deben corresponder a un solo concepto, representado por un término de indización; gramaticalmente puede tener la forma simple o compuesta, e esos casos son analizados para identificar si están de acuerdo con este principio.

*Ejemplos:*

	<i>(no conforme)</i>	<i>(sí)</i>
T2^a	606: ^aEducação de adultos	606: ^aAdultos^xEducação
T2^a	606: ^aDesenvolvimento curricular	606: ^aCurrículos^xEdesenvolvimento

- Los términos compuestos ‘Educação de adultos’ e ‘Desenvolvimento curricular’ representan más que uno concepto, e deberían ser desagregados para dar origen a encabezamientos subdivididos.

### T3, T4, T5

Los términos no coincidem com los índices, en cualquier campo o subcampo (nombres propios de persona, entidad, lugar).

*Ejemplos:*

	<i>(no conforme)</i>	<i>(sí)</i>
T3 <sup>a</sup>	600: ^aMaria Ulrich	600: ^aUlrich^bMaria^f1908-1988
T4 <sup>a</sup>	601: ^aEscola Superior de Educação de Portalegre	601: ^aInstituto Politécnico de Portalegre^bEscola Superior de Educação
T5 <sup>a</sup>	607: ^aAlqueva	607: ^aBarragem do Alqueva (Conselho de Portel, Portugal)

### **Encabezamiento específico**

En los Nombres Comunes, además de los tres ejes (UNIMARC, sintaxis, terminología) se estableció la variable del encabezamiento específico, para analizar la terminología al nivel de representación de conceptos, a través de la identificación de términos muy genéricos. El ‘principio de encabezamiento específico’ establece que “una materia debe ser representada por un encabezamiento cuyo sentido no sea ni más lato ni más restricto que esa materia”.

*Ejemplo: en el título:*

“Estudo da caracterização da qualidade ambiental na indústria corticeira em Portugal”,  
*indizado con el término ‘Agronomia’, es explícito que sería posible la atribución de un encabezamiento más específico:*

- Cortiça – Portugal
- Indústria da cortiça – Portugal
- Ambiente – Qualidade – Portugal

### Fuentes de terminología

Se confrontan los términos de indización en los registros, con los existentes en las fuentes de terminología habitualmente utilizadas: índices autorizados de la BNP y tesauros temáticos (Eurovoc, Agrovoc, DeCS- Descritores em Ciências da Saúde)

### **Análisis**

En las técnicas de análisis se utilizaran estadísticas descriptivas y para contrastar la dependencia de variables se utilizó el ‘coeficiente de correlación de *Pearson*’, que mide el grado de relación entre dos variables cuantitativas. Este coeficiente varía entre los valores:

-1 y 1, e los resultados pueden ser interpretados de la siguiente forma:

0.70 para más o para menos indica una fuerte correlación; 0.30 a 0.7, positivo o negativo, indica una correlación moderada; 0 a 0.30 indica una débil correlación.

### **Resultados**

Los resultados de nuestra investigación se presentan de acuerdo con los dos niveles de investigación:

#### **Nivel 1 – Instrumentos de apoyo à la indización en las BESP**

¿Cómo se procesa la indización en las bibliotecas de enseñanza politécnico en Portugal? (Respecto a utilización de normas, recurso a herramientas para recoger terminología, y utilización de fichero de autoridad). Los resultados obtenidos en la encuesta son:

En la variable **I1** (utilización de normas), la respuesta fue afirmativa en la mayoría de los casos (56,8%); el porcentaje de los que no utilizan normas es casi equiparada (43,2%).

De las respuestas afirmativas à la cuestión anterior, se puede verificar (tab.1) que cuando se pide la designación de tales instrumentos, 50% no responde, y 38,6% respondió de forma no especificada, es decir, las designaciones no se encuadran en los instrumentos considerados normativos, como NP 3715, NP 4036, o SIPORbase. La suma de 'não especifica' e de 'não responde' hace unos 88,6%, lo que es muy significativo, revelando que en la mayoría de los casos se hace la indicación sin observar la normativa.

<b>Valid</b>	<b>Frequency</b>	<b>Percent</b>
não especifica	17	38,6
NP 3715	2	4,5
Siporbase	2	4,5
NP 4036	1	2,3
não responde	22	50,0
<b>Total</b>	<b>44</b>	<b>100,0</b>

Tabla 1: Designación de los instrumentos normativos

En la variable **I2** (utilización de tesauros/ léxicos de materia), en la mayoría de los casos (81,8%) se usan, y en 18,2% no se usan.

Cómo es una cuestión abierta, se agruparon las respuestas de acuerdo con la tipología posible (tab.2). En algunos casos, en que se referirán más que dos herramientas, cómo no fueran designaciones uniformes, las categorizamos en 'varios'. En la categoría 'não especifica' se consideraran los casos en que las designaciones no se encuadran en respuestas válidas. Esta es la categoría con mayor número de respuestas (25%), e en unos 20,5 % no hay respuesta. En 22,7% de los casos se utilizan varias herramientas.

<b>Valid</b>	<b>Frequency</b>	<b>Percent</b>
não especifica	11	25,0
não responde	9	20,5
Eurovoc e TESE	1	2,3
Agrovoc	5	11,4
Lince	2	4,5
Decs	2	4,5
Unesco educação	1	2,3
Unesco e Lince	2	4,5
Mesh e Lince	1	2,3
vários	10	22,7
<b>Total</b>	<b>44</b>	<b>100,0</b>

Tabla 2: Designación de léxicos o tesauros

En la variable **I3** (utilización de fichero de autoridad), la respuesta fue equitativa para Sí y No (50%), la mitad de las bibliotecas no hace el control de autoridades.

## **Nivel 2 – Características de la indización en las BIPP**

### **Errores UNIMARC**

#### **Nombres Comunes (U1,U2)**

Las categorías U1 (errores de campo ^a) e U2 (errores de subcampos ^j, ^x, ^y, ^z) de esta variable, se observaran en los encabezamientos ‘no subdivididos’ y ‘subdivididos’; la U2 solo puede ocurrir en los encabezamientos de tipo ‘subdividido’.

Los errores U1 ocurren, con valores poco significativos en todas las bases, (excepto en B3TGE), (tab.3). Se encuentran errores de tipo U2 en las bases B1EDU, B3TGE e B4AGR, con un porcentaje inferior a 1% en la B2SAU y nula en la B5GER.

Los indicadores de subcampos en que ocurren más errores son en el ^x de subdivisión de materia, y en el ^j de subdivisión de forma. La no existencia de errores en subcampos ^y, ^z, se debe a la casi inexistencia de esos subcampos en los registros.

Cabeçalhos de Nomes Comuns					
UNIMARC	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
U1 (campo 606)	0,1%	2,3%	-	1,8%	1,2%
U2^j (subcampo)	-	-	0,7%	5,0%	-
U2^x (subcampo)	9,4%	0,8%	1,7%	2,5%	-

Tabla 3: Resultados de las variables UNIMARC (nombres comunes)

### Nombres Propios (U1,U3)

Los errores U1 de encabezamientos de nombres propios, se encuentran más en las bases B2SAU e B4AGR. Se trata en estos casos de nombres comunes que fueran registrados en los campos de nombres de persona e de lugares. También se detectaran campos vagos, que pueden ocurrir de un malo procedimiento, en la creación de registros.

Los errores de tipo U3 representan valores irrelevantes (tab.4), tratando-se del ^f (subdivisión de las fechas del nombre de persona).

Cabeçalhos de Nomes Próprios					
UNIMARC	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
U1 (campos)	50%	51,7%	4,1%	74,1%	-
U3 (subcampos)	-	1,7%	2,7%	0,6, %	-

Tabla 4: Resultados en las categorías de la variable UNIMARC (nombres propios)

### Errores de Sintaxis

#### **Nombres Comunes (S1)**

Los errores de sintaxis (orden de los subcampos) solo pueden ocurrir en los encabezamientos subdivididos (del total de encabezamientos, 26,3% son subdivididos. Los resultados de esta variable presentan el mayor porcentaje de errores en la base B1EDU (9,4%) y en la B4AGR (6,2%). En las bases B2SAU, B3TGE e B5GER ocurren menos errores.

Cómo es heterogéneo el número de encabezamientos subdivididos, era importante saber si la relación entre el número de errores y el número de encabezamientos subdivididos es significativa ¿ocurren más errores de sintaxis en los casos de más encabezamientos subdivididos?

Se aplicó el teste de correlación de *Pearson*<sup>4</sup> y la respuesta es:

En la situación A, en que se consideran todas las bases, la relación entre el número de encabezamientos e los errores de sintaxis es negativa, varía inversamente; hay más errores cuando el número de encabezamientos subdivididos decrece y vice-versa. No obstante este resultado puede ser influenciado por los valores de la base B1EDU, que representa un número mui elevado de errores, delante al menor número de encabezamientos subdivididos.

*correlación entre los encabezamientos subdivididos y los errores de sintaxis*

Situación A    -0,87

Situación B    0,01

En la situación B, después de excluir los valores de la B1EDU, la relación entre el número de encabezamientos y los errores de sintaxis es positiva, pero débil, lo que indica que el número de errores no depende linealmente del número de

---

<sup>4</sup>Coficiente positivo: implica la variación de variables en el mismo sentido; Coficiente negativo: implica la variación en sentidos opuestos; Coficiente de correlación superior a 0,7 indica relación (positiva o negativa) fuerte.

encabezamientos. Es posible que exista una relación no linear entre las variables, que sea relacionada con otros factores.

### **Nombres Propios (S2)**

En los encabezamientos de nombres propios no se encontraran errores de sintaxis, pues es mui escasa la ocurrencia de subdivisiones de materia.

### **Terminología**

#### **Nombres Comunes (T1, T2)**

Los términos verificados en los índices de materia de la BNP, y que no coinciden con esa terminología, los designamos de ‘no conforme’; en realidad no se trata de errores, pues a menudo los términos san recogidos de otras fuentes, como tesauros, más adecuados a temáticas específicas.

Verificamos que no es significativo el porcentaje de términos ‘no conforme’ en los subcampos  $\wedge_j$ ,  $\wedge_y$ ,  $\wedge_z$  (tab.5) de los encabezamientos subdivididos (que no ocurren con frecuencia, conforme ya referido en la variable UNIMARC).

<b>Cabeçalhos subdivididos</b>	<b>B1EDU</b>	<b>B2SAU</b>	<b>B3TGE</b>	<b>B4AGR</b>	<b>B5GER</b>
T1 $\wedge_a$ (assunto principal)	12,9%	6,5%	7,3%	53,1%	8,4%
T1 $\wedge_j$ (subdiv. forma)	0,5%	-	4,0%	4,3%	-
T1 $\wedge_x$ (subdiv. assunto)	3,7%	11,3%	8,3%	51,8%	7,8%
T1 $\wedge_y$ (subdiv. lugar)	-	-	2,1%	4,3%	3,6%
T1 $\wedge_z$ (subdiv. cronológica)	-	-	-	-	1,2%

Tabla 5: Terminología ‘no conforme’ los índices BNP (encabezamientos subdivididos)

En los encabezamientos ‘no subdivididos’ se verifica la terminología observando igualmente los términos simples y los compuestos, correspondiendo a las categorías T1 y T2 respectivamente (tab.6).

<b>Cabeçalhos Não subdivididos</b>	<b>B1EDU</b>	<b>B2SAU</b>	<b>B3TGE</b>	<b>B4AGR</b>	<b>B5GER</b>
T1 ^a (termos simples)	5,1%	14,1%	7,7%	8,7%	0,6%
T2 ^a (termos compostos)	25,0%	30,3%	32,1%	15,6%	4,2%

Tabla 6: Terminología ‘no conforme’ los índices BNP (encabezamientos No Subdivididos)

La suma de las ocurrencias en ^a de los encabezamientos ‘no subdivididos’ e de los ‘subdivididos’, y la ocurrencia en la subdivisión ^x en estos últimos, nos da la globalidad de términos ‘no conforme’ (tab.7).

<b>Totais</b>	<b>B1EDU</b>	<b>B2SAU</b>	<b>B3TGE</b>	<b>B4AGR</b>	<b>B5GER</b>
T1^a/T2^a (assunto principal)	43, 0%	50,9%	47,1%	77,4%	13,2%
T1^x (subdivisão de assunto)	3,7%	11,3%	8,3%	51,8%	7,8%

Tabla 7: Total de términos ‘no conforme’ (encabezamientos NS y Subdivididos).

El porcentaje de términos ‘no conforme’ es mayor en la base **B4AGR**. Esta base tiene una temática específica, y sería de esperar que en las bases **B1EDU**, **B2SAU** e **B3TGE**, con la misma característica, los valores fueran semejantes, pero son más bajos.

Genéricamente, hay más términos ‘no conforme’ en la materia principal, lo que está de acuerdo con el facto de haber más encabezamientos ‘no subdivididos’ en las tres primeras. En el caso de la **B4AGR**, que tiene mayor número de encabezamientos ‘subdivididos’ no se podrá establecer esa relación. La base **B5GER** es la que presenta mayor conformidad con los índices de la BNP.

### Nombres Propios (T3, T4, T5)

La distribución de términos ‘no conforme’ en la variable de nombres propios es muy heterogénea (tab.8).

<b>Nomes Próprios</b>		B1	B2	B3	B4	B5
<i>(Número de cabeçalhos)</i>		<i>(4)</i>	<i>(58)</i>	<i>(620)</i>	<i>(472)</i>	<i>(128)</i>
<i>UNIMARC</i>	<i>U1 (campo)</i>	50%	51,7%	4,1%	74,1%	-
	<i>U2 (subcampos)</i>	-	1,7%	2,7%	0,6%	-
<i>Terminologia</i>	<i>T3 (pessoa)</i>	25%	-	0,2%	-	0,7%
	<i>T4 (entidade)</i>	-	-	35,4%	0,2%	0,7%
	<i>T5 (geográfico)</i>	-	3,4%	0,3%	20,7%	-

Tabla 8: Distribución de errores en las variables de Nombres Propios

Sintetizando los resultados para una observación general, verificamos que en los encabezamientos de nombres comunes (tab.9) los porcentajes más significativos se presentan en la variable ‘Terminología’; la más elevada en la B4AGR (total 77,4%), y la más baja en la B5GER (13,2%).

En la variable UNIMARC, el total de las categorías U1 y U2, la ocurrencia de errores es poco significativa en general.

<i>Nomes Comuns</i>	B1EDU	B2SAU	B3TGE	B4AGR	B5GER
UNIMARC (U1,U2)	0,1%	2,1%	8,7%	12,9%	1,2%
Sintaxe	9,4%	0,6%	1,9%	6,2%	1,8%
Terminologia (T1,T2)	43,0%	50,9%	47,1%	77,4%	13,2%
<i>Nomes Próprios</i>					
UNIMARC (U1, U3)	50,0%	53,4%	6,8%	74,7%	-
Terminologia (T3,T4,T5)	25,0%	3,4%	35,9%	20,9%	1,4%

Tabla 9: Suma de los resultados por variable en los encabezamientos de NC e NP

En los encabezamientos de nombres propios, es la variable UNIMARC la que presenta mayor número de errores.

### **Encabezamiento Específico**

En esta variable es considerable el número de registros en los que no se cumple el ‘principio de encabezamiento específico’. En las bases B1EDU (83,1%) y B4AGR (62,5%) se observan encabezamientos muy genéricos, lo que resulta curioso, pues se tratan de áreas temáticas específicas. En las bases B2SAU e B3TGE, el porcentaje de errores es respectivamente de 41,1% y 27,4%.

Para saber si hay relación entre la variable de ‘encabezamiento específico’ y los encabezamientos ‘subdivididos’ y ‘no subdivididos’ se aplicó el teste de correlación de *Pearson*:

- existe una relación positiva entre los encabezamientos ‘no subdivididos’ y los errores de ‘encabezamiento específico’; la ocurrencia de estos errores tiende a crecer con el aumento de encabezamientos ‘no subdivididos’.
- Existe una relación negativa entre los encabezamientos ‘subdivididos’ e los errores de ‘encabezamiento específico’; la ocurrencia de errores en estos tiende a decrecer con el aumento de los encabezamientos subdivididos.

### **Fuentes de terminología**

La mayor cobertura de índices de la BNP se verifica en la B5GER, pues se trata de un fondo documental de carácter más general, donde las temáticas dominantes se encuentran en esos índices de materia. En las bases B2SAU e B4AGR se verifica la menor cobertura, una vez que son bases con temáticas específicas en las áreas de la salud y de ciencias agrarias, que recurren más a los tesauros.

En todos los casos, excepto de B5GER, se recurre más a los tesauros para la recogida de términos de indización, en concreto de Eurovoc, la B1EDU (65,3%) y B3TGE (54,2%), de Agrovoc en la B4AGR (28%) y de la lista de descriptores DeCs en B2SAU (54,5%).

Sintetizando los resultados de las últimas variables (tab.10), es curioso que en la B1EDU que tiene un porcentaje de errores tan elevado en la variable ‘encabezamiento específico’, sean utilizados términos de los tesauros (65,3%), pues los tesauros son, en principio, más específicos. Cómo en este caso se trata del tesoro Eurovoc, se explica, porque no es un tesoro específico de educación. En el caso de la B4AGR, a pesar de ser utilizado el tesoro Agrovoc, que es adecuado a las temáticas, hay muchos términos que no coinciden con sus descriptores.

Los porcentajes señalados a gris no suman los 100% porque, en este caso, se identificaran términos iguales en las dos fuentes de terminología (se trata de términos muy genéricos).

<b>Cabeçalho específico</b>		B1EDU	B2SAU	B3TGE	B4AGR	B5GER
		83,1%	41,1%	27,4%	62,5%	6,0%
<b>Fontes de Terminologia</b>	Índices BNP	43,5%	38,5%	42,5%	24,5%	76,6%
	Tesouro/Lista	65,3%	54,5%	54,2%	28,0%	21,1%

Tabla 10: Variables ‘Encabezamiento específico’ y ‘Fuentes de Terminología’

## Conclusiones

De la investigación en las bibliotecas politécnicas se concluye que, en la mayoría de ellas se practica la indización sin orientación normativa, lo que consideramos sorprendente, una vez que el conocimiento normativo debe ser la base de todo el proceso. Se verifica alguna confusión en las respuestas, respecto a la designación de instrumentos normativos y de herramientas de apoyo a la indización. En la mitad de estas instituciones no se utiliza un fichero de autoridad de materias.

En el grupo de bibliotecas en que se observaran características de los índices de materia, se concluye que:

- El porcentaje de errores en el formato UNIMARC no es significativo, la media de errores es de 3%. No obstante, se debe considerar esta característica, una vez que la vinculación entre el catálogo bibliográfico y el catálogo de autoridades se hace a través de las etiquetas UNIMARC;
- Respecto a la sintaxis de los encabezamientos, el porcentaje de errores tan poco es significativa, y el predominio de encabezamientos ‘no subdivididos’ tiene ahí influencia;
- Cuanto a la terminología, hay procedimientos de indización diferentes en las bibliotecas; se utiliza el lenguaje natural, descriptores de variadas fuentes de información, o encabezamientos pre-coordinados;
- La recogida de términos se hace esencialmente a partir de tesauros, y no se practica el control de autoridades, tanto a nivel de la terminología como de la sintaxis de los encabezamientos;

Los resultados de este nivel evidenciaran la dificultad en el proceso de indización en sus principios esenciales, como la identificación de conceptos y su representación de forma controlada. A pesar de los datos obtenidos no permitieren identificar el tipo de dificultades por parte de los profesionales, podremos afirmar que algunos errores no ocurrirían si la normativa fuera observada, como en el caso del ‘encabezamiento específico’.

Podemos deducir de estos resultados, que lo tipo de problemas detectados en los índices de materia, podrán ocurrir igualmente en las bibliotecas politécnicas, pues estas reflejaran dificultades en el proceso de indización.

Las cuestiones que se colocaran en el primer nivel de investigación (en las bibliotecas politécnicas), dan respuestas cuanto a la utilización de normas y herramientas de apoyo a la indización. El análisis de los índices de materia en el segundo nivel (estudio de caso) permite relacionar las características e calidad de la

indización con una inadecuada utilización normativa y con la falta de control terminológico.

Así, los principales puntos de convergencia entre los dos niveles son (tab.11): el indicio de que existe una débil utilización normativa (11%), se manifiesta en las características de los índices del grupo de bibliotecas (no son observados principios básicos de indización). La variedad de fuentes de terminología (23% utiliza más que dos), y la reducida utilización del fichero de autoridades (50%), potencia la terminología no controlada.

Indícios nas BESP (nível 1)		Manifestações nas BIPP (nível 2)
Utilização de Instrumentos normativos	43% não usam	- Não observação de princípios de indexação, (cabeçalho específico e uniforme)
	11% normativos	
	4,5% SIPORbase	- Terminologia não conforme o SIPORbase
Várias fontes de terminologia	32% até 2 tesouros	- Utilização indiscriminada de descritores, palavras-chave, cabeçalhos pré-coordenados
	23% mais que 2	
Ficheiro de autoridades	50% não utilizam	- Terminologia não controlada

Tabla 11: Relación de los dos niveles de investigación: los indicios y las manifestaciones

Cuanto a la utilización del sistema SIPORbase, se concluye que es muy poco utilizado en las bibliotecas politécnicas (4,5%), y solo como instrumento normativo, pues los índices de la BNP no son referidos como una herramienta de apoyo.

Tales constataciones refuerzan la línea central de nuestra investigación, de que es necesario orientar el proceso de indización, e de que la utilización de un vocabulario controlado ayuda a la consistencia de los índices de materia. La falta de consistencia observada en los índices demuestra la necesidad de una política de indización, que permita orientar el trabajo de los profesionales. Con esa finalidad, y culminando nuestra investigación, se ha presentado un modelo que se sugiere utilizar.

## **Propuesta de Modelo de Indización**

Se propone un modelo para desarrollar el proceso de indización que representa las tres etapas básicas del proceso. Le atribuimos el nombre de IRNA, acrónimo de Identificación, Recogida, Normalización e Autoridades.

1 – Identificación y selección de conceptos (análisis del documento para definición de su contenido, según la NP 3715; relación con los FRBR y categorías de la norma ISO 25964-1.

2 - Recogida de términos de fuentes de terminología, de acuerdo con recursos recomendados (Índices BNP / EUROVOC / UNESCO Thesaurus / DeCS – Descritores em Ciências da Saúde / AGROVOC / UNESCO Nomenclature / GEMET – General Multilingual Environmental / VIAF).

3 - Normalización de términos y de encabezamientos, y respectivo control de autoridades. (Recomendaciones para establecer las relaciones semánticas posibles).

Al final de la descripción de las tres etapas en el proceso de indización, se presenta un diagrama de flujo del modelo IRNA.