

Indización asistida por computadora: una experiencia en el área nuclear

*Mercedes Bavio (1), Alejandra T. Chávez Flores (1),
Gabriela Puglia (1) y Marisa Velazco Aldao (2)*

- (1) *Comisión Nacional de Energía Atómica (CNEA). Centro de Información: Biblioteca “Eduardo J. Savino” Centro Atómico Constituyentes. Avda. General Paz 1499, 1650, San Martín, Prov. Buenos Aires. República Argentina.*
- (2) *Comisión Nacional de Energía Atómica. Biblioteca “Leo Falicov”, Instituto Balseiro. Centro Atómico Bariloche, Av. E. Bustillo km 9,500, 8400 Bariloche. Prov. de Río Negro. República Argentina.*

E-mails: bavio@cnea.gov.ar; chavez@cnea.gov.ar; puglia@cnea.gov.ar; velazcom@cab.cnea.gov.ar

RESUMEN

Se presenta un estudio de caso sobre la evaluación preliminar que se está llevando a cabo en el Centro de Información CAC y en la Biblioteca “Leo Falicov” de la Comisión Nacional de Energía Atómica, en el uso de un sistema de indización asistida por computadora CAI¹ aplicado a la asignación de descriptores a los registros bibliográficos para su envío al INIS². Se describen las funcionalidades de la interfaz de dicho sistema, los aspectos metodológicos del proceso de indización bajo esta nueva modalidad y una evaluación inicial de ventajas y problemas encontrados. Durante las primeras pruebas, el sistema CAI ha demostrado ser una herramienta de gran ayuda para la tarea del indizador, al ampliar el espectro de términos a asignar de manera amigable, ágil y flexible, contribuyendo a la calidad de la indización. Se evidencia además que, aún en ciertas disciplinas muy específicas, se hace necesario que la indización por computadora sea complementada con la intervención de profesionales expertos en las temáticas consideradas.

PALABRAS CLAVE

CAI; CNEA; INDIZACIÓN ASISTIDA POR COMPUTADORA; INIS; OIEA; RECUPERACIÓN DE LA INFORMACIÓN

INTRODUCCIÓN

Este trabajo tiene como objetivo presentar la evaluación preliminar que se está llevando a cabo en el Centro de Información CAC y en la Biblioteca “Leo Falicov” de la Comisión Nacional de Energía Atómica utilizando un sistema de indización asistida por computadora aplicado a la asignación de descriptores a los registros bibliográficos a ser enviados a INIS.

INIS es el principal sistema mundial de información en el área de los usos pacíficos de la energía nuclear. Dicho sistema está gestionado por el Organismo Internacional de Energía Atómica³ a través de la Secretaría de INIS, en colaboración con sus Estados Miembros y otras organizaciones internacionales.

¹ *Computer-assisted Indexing.*

² *International Nuclear Information Systems.*

³ El OIEA es un organismo autónomo que forma parte del sistema de las Naciones Unidas.

Nuestro país es miembro de INIS desde sus inicios, en el año 1970. Una de sus principales responsabilidades como Centro Cooperantel es el envío de los registros bibliográficos y el texto completo (en aquellos casos en que los derechos de autor lo permitan) de toda la producción intelectual publicada en nuestro país dentro del alcance temático de INIS. Esta actividad es centralizada en las unidades de información de CNEA.

A partir de la información enviada por cada Estado Miembro, se mantiene actualizada la base de datos INIS cuyo acceso, a partir de noviembre de 2008, es libre y gratuito para toda la República Argentina, en la siguiente dirección: <http://inisdb2.iaea.org>.

La calidad de la base de datos INIS está determinada por reglas y normas de ingreso, formatos, códigos y listas de autoridades contenidas en los Manuales de Referencia que el sistema dispone. Estos manuales son herramientas esenciales para los catalogadores, indizadores, y otros especialistas de información que participan en INIS.

Para el procesamiento de los registros bibliográficos, INIS dispone de un *software* denominado WinFIBRE⁴, cuyas principales prestaciones son las siguientes:

- Facilita el ingreso de toda la información bibliográfica que INIS necesita para cada tipo de documento.
- Controla y añade términos y abreviaturas del tesoro, de las listas de autoridad corporativa, códigos de categoría temáticas, títulos normalizados de publicaciones periódicas y otras autoridades.
- Controla y verifica que el ingreso de los datos de los registros se ajuste a las reglas y normas de INIS.
- Genera un archivo para el envío de los registros procesados a la Secretaría de INIS.

Dado que una de las tareas esenciales, para una eficiente y efectiva recuperación de la información, es la correcta indización, la Secretaría de INIS ha venido poniendo esfuerzos en encontrar las herramientas tecnológicas que faciliten dicha tarea y al mismo tiempo disminuyan los costos involucrados.

A tal fin, se creó un grupo de trabajo formado por los Oficiales de Enlace INIS de los siguientes países: Australia, Argentina, Belarús, China, Francia, Alemania, Japón, Nueva Zelanda, Federación Rusa, Suiza, Turquía y Estados Unidos de América. Su misión fue asistir a la Secretaría en los siguientes asuntos:

- Evaluar cómo la indización asistida por computadora podía ayudar a los Centros Cooperantes encargados de procesar la información y documentación a ser enviada a la base de datos.

⁴ *Windows Friendly Input of Bibliographic Records*

- Identificar, probar y seleccionar potenciales tecnologías y *softwares* que se adecuaran a las necesidades de INIS.

Como resultado del trabajo de dicho grupo, se seleccionó una herramienta desarrollada por la firma Convera™ que cumplía con los requerimientos establecidos.

De esta manera la Secretaría INIS ha podido ofrecer a los Centros Cooperantes una aplicación que permite:

- Mantener la calidad de la base de datos INIS.
- Mejorar el análisis temático realizado manualmente.
- Optimizar la consistencia de la indización.
- Reducir el tiempo de la catalogación mejorando el aprovechamiento de los recursos humanos disponibles.

PROCESO DE INDIZACIÓN

El proceso de asignación de descriptores, en el INIS, puede llevarse a cabo de las siguientes maneras:

1. Proceso Estándar (la indización la realiza el indizador en forma manual utilizando el Tesauro incluido en WinFIBRE).
2. Proceso CAI *on batch* en el que no se utiliza una interfaz de interacción con el *software* de Convera™.
3. Proceso CAI *on line*, proceso interactivo en línea, accediendo al servidor de INIS en Viena.

En esta ponencia se describe este último proceso.

Para comenzar a trabajar es imprescindible cargar previamente en el sistema CAI instalado en el servidor del OIEA en Viena, el archivo resultante del procesamiento (catalogación e inclusión del resumen) de los registros bibliográficos procesados con WinFIBRE. Se debe tener presente que el CAI es un módulo sólo para la indización y que trabaja teniendo en cuenta (por el momento) sólo el título y el resumen en inglés.

A continuación se presenta una pantalla de trabajo del sistema CAI *on line* (Fig. 1):

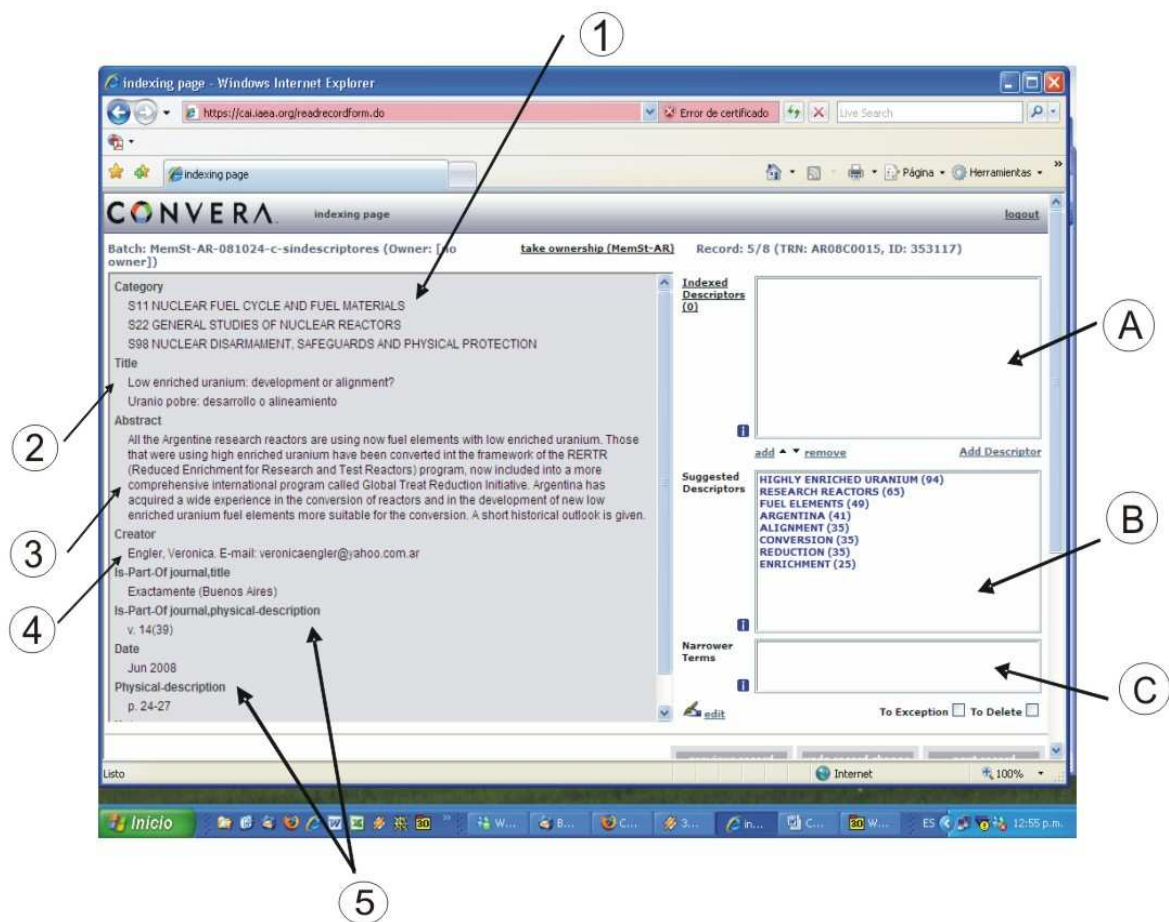


Figura 1

En la parte izquierda de esta pantalla se observa el contenido de un registro seleccionado

1. Categoría temática.
2. Título del trabajo en inglés y en castellano.
3. Resumen en inglés.
4. Autor/es personales o institucionales.
5. Determinados campos bibliográficos seleccionados.

La zona derecha de la pantalla es el área de trabajo con los descriptores. El sistema muestra allí tres ventanas:

- A. la superior, que inicialmente se presenta vacía, donde se alojarán todos los descriptores que resulten seleccionados una vez finalizado el proceso de indexación.
- B. la del medio, que contiene la lista de todos los descriptores sugeridos por el sistema CAI. Al hacer la selección de los términos, el sistema automáticamente elimina aquellos que son más generales (BT o *broader term*), los que serán automáticamente asignados en un proceso posterior externo al CAI.
- C. la inferior, en donde cada vez que se seleccione un descriptor en la ventana del medio, se mostrarán los términos más específicos que corresponden al mismo.

La metodología a seguir comprende los siguientes pasos:

- Antes de seleccionar un término sugerido por el CAI, se debe verificar si realmente es apropiado para el registro en proceso. A tal fin el sistema permite hacer clic sobre el término sugerido y ver su ubicación (ya sea en el título y/o resumen) que se visualiza en forma resaltada y en color (puede ser el mismo término o un sinónimo).
- En aquellos casos en que el descriptor sugerido por CAI en la segunda pantalla tenga descriptores específicos (NT o *narrow term*), el sistema los visualizará en la ventana inferior. Por lo tanto, se debe verificar si alguno de ellos es más apropiado y en ese caso se lo seleccionará haciendo clic en *Add*. Esta acción hará que dicho descriptor se traslade automáticamente a la ventana superior.
- Se deben agregar todos aquellos descriptores que se consideren adecuados para el contenido del documento seleccionándolos ya sea del grupo de los sugeridos por CAI o del grupo de los específicos para cada caso, haciendo clic en *Add* para que los mismos sean trasladados a la ventana superior.
- Los descriptores sugeridos que resulten muy generales, no relevantes o no apropiados para el contenido del documento, no se seleccionan y se pasan por alto, pero permanecen en su ventana correspondiente. Dichos descriptores posteriormente desaparecerán en el proceso de exportación de los registros al sistema que realiza el control final de calidad antes de enviarlos a la base de datos INIS.

En aquellos casos en que sea necesario agregar un término no sugerido inicialmente por el sistema, existe la opción *Add descriptor* a través de la cual se accede al Tesauro INIS en una ventana que se abre por separado. Ello permite:

- Agregar descriptores no sugeridos por el CAI, pero que son relevantes para el contenido de un documento en particular, por ejemplo:
 1. Descriptores para asignar valores numéricos: valores de temperatura, presión, energía, etc.
 2. Descriptores para reacciones nucleares.
 3. Descriptores para componentes químicos, aleaciones, etc.
 4. Descriptores que se refieran a aspectos históricos.
 5. Descriptores que designa determinado tipo de material bibliográfico: conferencias, informes, etc.
- Visualizar la relación entre el descriptor que se está considerando en el Tesauro y aquellos que ya están seleccionados (Fig. 2). Esto es posible porque, de corresponder, el CAI coloca un botón de color que responde a los siguientes conceptos:
 - **Rojo oscuro:** descriptores ya indizados (asignados y alojados en la primera ventana)
 - **Rojo suave:** término general (BT) de un descriptor sugerido o indizado.

- **Amarillo:** término específico (NT) de un descriptor sugerido o indizado.
- **Verde:** descriptor ya sugerido.

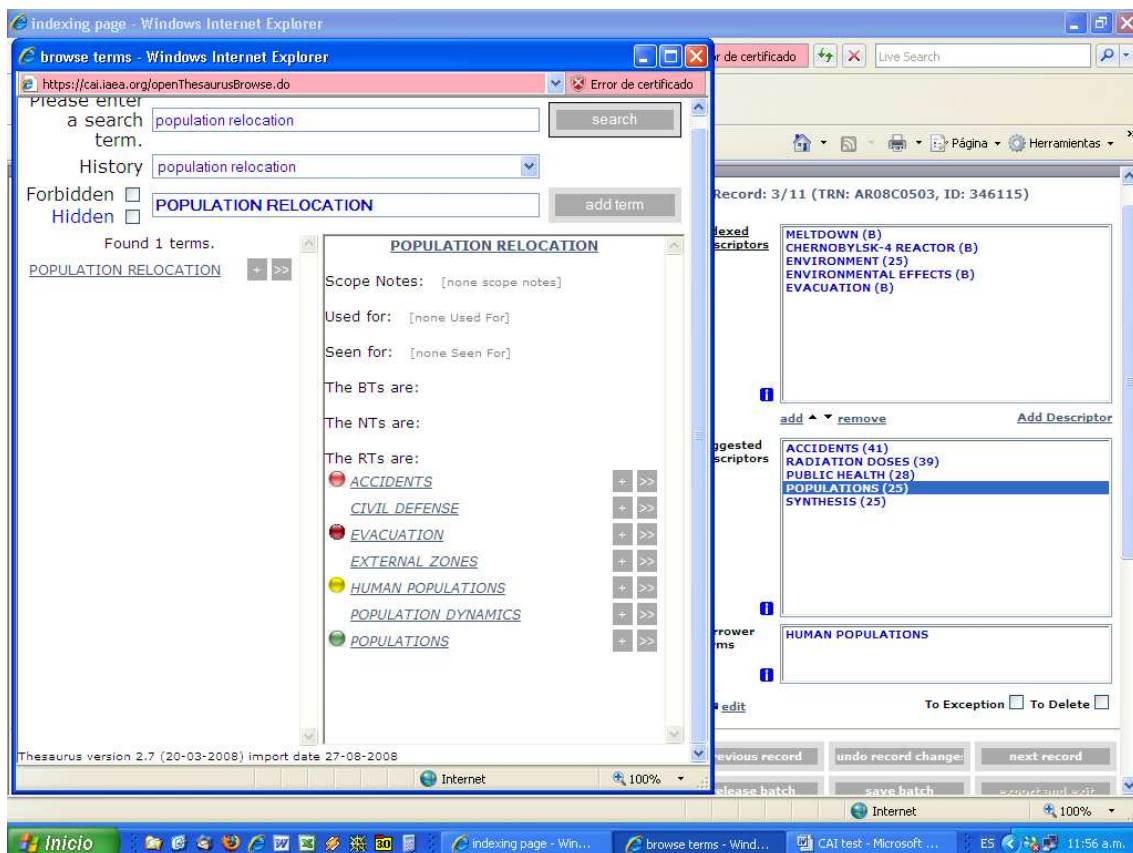


Figura 2

Una vez que se considera que han sido asignados todos los descriptores pertinentes, se graba el registro y a continuación el sistema presenta el registro bibliográfico siguiente para su procesamiento, operación que se repite hasta llegar al último registro contenido en el archivo con el que se está trabajando. Una vez procesados todos los registros, el archivo resultante es enviado al sistema de producción regular de INIS para otros procesos de control de calidad que se efectúan antes de incluir los registros procesados por CAI en la base de datos.

VENTAJAS Y PROBLEMAS

A partir de los resultados de las primeras pruebas con CAI *online* se está en condiciones de realizar los siguientes comentarios iniciales:

Ventajas:

- ✓ El CAI provee una herramienta muy útil y práctica posibilitando la detección, y por lo tanto la selección, de descriptores que no hubieran sido asignados en una indización manual.

- ✓ Tanto para los casos de fórmulas matemáticas y químicas como para determinadas abreviaturas, el sistema los capta y automáticamente adjudica los descriptores pertinentes.
- ✓ Permite, asimismo, proponer manualmente descriptores nuevos, no incluidos en el Tesauro INIS, los que una vez evaluados y aceptados por los especialistas de INIS pasan a formar parte de dicho Tesauro.
- ✓ Presenta una interfaz gráfica para usuario muy amigable y flexible.
- ✓ La combinación entre el sistema de indización asistida y la intervención humana resulta muy enriquecedora en la práctica.

Al respecto cabe mencionar que, inicialmente, tanto el personal bibliotecario como el especialista en temas nucleares que colabora con la indización, tuvieron una actitud reticente y hasta escéptica respecto de las ventajas que el sistema estaba en condiciones de brindar. Ello se debió a que ya se había trabajado con la versión del CAI *off line*, donde los registros se procesan sin la intermediación de una interfaz y no había resultado amigable. A título de ejemplo, no era sencillo detectar a partir de qué término el sistema sugería los descriptores. Después de la primera prueba utilizando la versión CAI *on line*, esa apreciación cambió radicalmente y el sistema empezó a verse como una herramienta atractiva por su amigabilidad, agilidad y flexibilidad.

Problemas:

- Los registros deben constar indefectiblemente con un resumen y el mismo debe estar bien redactado en inglés. De otra manera, el CAI sugerirá descriptores inadecuados. Este es uno de los puntos más problemáticos pues, muchas veces los trabajos no incluyen el resumen en inglés que debe ser por lo tanto realizado en el Centro Cooperante, lo que implica un alto costo en tiempo y en recursos humanos capacitados. Cabe destacar que, por regla general, al no ser el inglés lengua materna entre los investigadores, se detectan resúmenes con redacción de dudosa calidad.
- Si el resumen existe pero es muy reducido, el sistema no podrá ser exhaustivo en la cantidad de descriptores propuestos. (INIS aconseja que cada registro contenga no menos de 5 descriptores).
- Una problemática importante se presenta cuando un concepto aparece de diferentes maneras (*differrent appearances*) y el CAI le asigna uno o más descriptores. Estos casos son similares a los términos no permitidos de un tesauro (FT o *forbiden term*) y el uso de una o más relaciones de “véase”. Para mejorar el procesamiento de estos casos, los especialistas de INIS mantienen un repositorio de los denominados “términos ocultos” (*hidden terms*). Estos términos no son incorporados al Tesauro, aunque son considerados como una extensión del mismo.

Ejemplos de casos considerados:⁵

⁵ Ejemplos provistos por el Sr. A. Nevyjel de INIS

Compuestos:

Descriptor	término oculto	texto libre
MAGNESIUM BORIDES	MgB_2	MgB ₂
MAGNESIUM CARBONATES	MgCO_3	MgCO ₃
ARSENIC IONS	As"3"-	As ³⁻

Isótopos:

Descriptor	término oculto	texto libre
CESIUM 137	"1"3"7cs 137 caesium caesium 137 137 cesium	¹³⁷ Cs 137 Caesium, 137-Caesium Caesium 137, Caesium-137 137 Cesium, 137-Cesium

Partículas elementales:

Descriptor	término oculto	texto libre
B QUARKS	bottom quarks	
T QUARKS	top quarks	
ELECTRON NEUTRINOS	#nu#_e	ν _e

Términos UK/US

Descriptor	término oculto
A CENTERS	a centres
ANALOG COMPUTERS	analogue computers
ANALOG SYSTEMS	analogue systems
ANESTHESIA	anaesthesia

Signos diacríticos y nombres de países

Descriptor	término oculto
Signos diacríticos:	
BAECKLUND TRANSFORMATION	backlund transformation
BRUECKNER METHOD	bruckner method
MOESBBAUER EFFECT	mossbauer effect

Nombres de países:

CAMBODIA	kampuchea
GREECE	hellas
MYANMAR	burma

Otros casos:

Descriptor hidden term

Singular/Plural:

FUNGI	fungus
FUNGI	funguses
G MATRIX	g matrices
G MATRIX	g matrixes

Reverse Sequence:

ATOM-MOLECULE COLLISIONS	molecule-atom collisions
ATOM-MOLECULE COLLISIONS	atom-molecule scattering
ATOM-MOLECULE COLLISIONS	molecule-atom scattering

Variaciones gramaticales:

PERIODICITY	periodic
PERIODICITY	periodical
PERIODICITY	periodically

Frases vs. términos compuestos:

RADIOWAVE RADIATION	radio wave
SPACE-TIME	spacetime
WAVE FUNCTIONS	wavefunction

Terminología:

GAMMA SPECTROMETERS	#gamma#ray spectrometer
GAMMA SPECTROMETERS	#gamma#-ray spectrometer
GAMMA SPECTROMETERS	gammaray spectrometer
GAMMA SPECTROMETERS	gamma-ray spectrometer

Abreviaturas:

CARBON DIOXIDE LASERS	CO_2 laser
CARBON DIOXIDE LASERS	CO2 laser
KOBAYASHI-MASKAWA MATRIX	CKM matrix

Valores numéricos:

KEV RANGE	kev
MEV RANGE	mev
GEV RANGE	gev

En la actualidad los términos ocultos suman unos 64.400

En algunos casos es muy evidente que la intervención humana es indispensable para dirimir si un descriptor debe ser asignado o no. Los términos que se prestan a ello, por ejemplo, son los que tienen más de un significado:

MPA

- MAXIMUM PERMISSIBLE ACTIVITY
- Mega Pascal (MPa)

GDP

- GROSS DOMESTIC PRODUCT
- GADOLINIUM PHOSPHIDES (gDp)

Otros candidatos a estar erróneamente asignados son los siguientes:

PRODUCTION

DECAY

TRANSPORT

CONSIDERACIONES FINALES

A partir de las pruebas que se han realizado con el CAI, es posible comentar que se trata de una herramienta sumamente útil para el trabajo de indización. La tecnología subyacente combina técnicas de recuperación de información tradicional con otras modernas y avanzadas, que posibilitan el entrenamiento del sistema para mejorar su rendimiento. Sin embargo, se encuentra que en áreas tan específicas como la nuclear, la intervención humana sigue siendo aún necesaria. Dicha situación continuará hasta tanto se llegue a una sofisticación tal que permita alcanzar niveles de entrenamiento de los sistemas de indización, para que los mismos se puedan utilizar en forma totalmente automática sin provocar asignación de descriptores en forma errónea. La posibilidad que brinda el INIS a los Centros Cooperantes de utilizar el CAI en forma remota accediendo a su servidor ubicado en Viena ha permitido a las autoras ganar experiencia en la utilización de este tipo de sistemas. Se hace necesario continuar con las pruebas y, sobre todo, con el intercambio de experiencias con otros Centros Cooperantes a fin de ir mejorando su uso en forma progresiva.

AGRADECIMIENTOS

Las autoras agradecen al Sr. Alexander Nevyjel, *Group Leader, Content Management Group de INIS* por permitirles incluir ejemplos contenidos en presentaciones de su autoría; al Dr. Renato Radicella por su incondicional y desinteresada colaboración en el trabajo de indización para INIS y a la Lic. Alicia Cassanello, Jefa del Sector Tratamiento Documental del Centro de Información CAC, por sus aportes en la elaboración del presente documento.

BIBLIOGRAFIA

Bosch, M. (2006). La indización para el manejo de conocimiento en organizaciones frente al tratamiento documental en bibliotecas. *Biblios: Revista de Bibliotecología y Ciencia de la Información*, 7 (23), 1-11.

Gil Leiva I., Rodríguez Muños J. V. (1996). Tendencias en los sistemas de indización automática: estudio evolutivo. *Revista Española de Documentación Científica*, 19 (3), 273-291.

INIS: Diccionario Multilingüe, Parte V: español-inglés-alemán-francés-ruso. (1995). *IAEA-INIS-20 (Rev. 1) (S)*.

INIS: FIBRE User's Manual. (1999). *IAEA-INIS*.

Nevyjel, A. (2003). Computer assisted indexing functionality concept, software evaluation and implementation plan. *Salzburg: ODOK'03. International Atomic Energy Agency*. (Power Point).

Nevyjel, A. (2006). Computer-assisted indexing for the INIS database. *IAEA: Information and Innovations: Views, Trends, Prospects*, (3), 15-20. Disponible en: http://www.iaea.org/inisnkm/docs/article_cai.pdf
Consultado: 06/11/08.

OIEA. (2002). *Presentación del INIS. GEN/PUB/013 (Rev. 12)*.

Rodríguez Perojo K., Ronda León R. (2006). Organización y de la información: un enfoque desde la perspectiva de la automatización. *ACIMED*, 14 (1). Disponible en: http://bvs.sld.cu/revistas/aci/vol2_2_94/aci04106.htm
Consultado: 06/11/08.

Subject Categories and Scope Descriptions. (2002). *ETDE/INIS Joint Reference Series N° 2*.