

ANÁLISIS DEL CONTENIDO TEMÁTICO DE LA REVISTA NUCLEUS

Ramiro Guerra Valdés

Centro de Gestión de la Información y Desarrollo de la Energía (CUBAENERGÍA)

Calle 20 N° 4111 e/ 18-A y 47, Miramar, Playa, Ciudad de La Habana, Cuba

rguerra@cubaenergia.cu

Resumen

Se presentó un programa de computación para realizar análisis estandarizado de áreas de investigación y conceptos claves de la ciencia y la tecnología nucleares desarrollado en CUBAENERGÍA. Se describieron los componentes principales del sistema de procesamiento de información, así como los métodos computacionales y módulos para el análisis del contenido temático de archivos de registros de la base de datos del Sistema Internacional de Información Nuclear. Se reportan los resultados del análisis del contenido temático de la revista Nucleus desde 1986 hasta 2005. Se mostraron los resultados del estudio demostrativo Nucleus, Ciencia, Tecnología y Sociedad. Estos resultados aportan nuevos elementos para valorar el significado del contenido de la revista Nucleus en el contexto de la innovación en áreas de investigación multidisciplinarias interrelacionadas.

ANALYSIS OF THE THEMATIC CONTENT OF REVIEW NUCLEUS

Abstract

A computer program for performing standardized analysis of research areas and key concepts of nuclear science and technology under development at Cubaenergia is presented. Main components of the information processing system, as well as computational methods and modules for thematic content analysis of the International Nuclear Information System Database record files are described. Results of thematic content analysis of journal Nucleus from 1986 to 2005 are shown. Furthermore, results of demonstrative study Nucleus, Science, Technology and Society are also shown. The results provide new elements to assess the significance of the thematic content of journal Nucleus in the context of innovation in interrelated multidisciplinary research areas.

Key words: information retrieval, information systems, INIS, knowledge base, mathematical models, standardized terminology

1. INTRODUCCIÓN

La revista Nucleus en el transcurso de los años se ha establecido como la principal publicación para difundir la actividad científico-técnica del país en el ámbito de la ciencia y la tecnología nucleares. Todos los artículos científicos publicados en Nucleus están registrados en la base de datos del Sistema Internacional de Información Nuclear (INIS). Este hecho, que apenas llama la atención de los lectores, es el elemento clave que permite analizar el contenido temático de la revista desde la perspectiva del estado actual de la ciencia y tecnología nucleares a nivel mundial.

La base de datos INIS es el resultado de la cooperación científico-técnica internacional en el marco del Organismo Internacional de Energía Atómica (OIEA) y actualmente contiene cerca de 2,6 millones de registros de documentos científico-técnicos, con un crecimiento anual promedio de 60 000 a 80 000 nuevos registros. El servidor INIS Cuba instalado en CUBAENERGÍA,

facilita el acceso a los registros de datos bibliográficos y proporciona la información a los usuarios. Para obtener mayor provecho de esta inmensa fuente de información de alta calidad, se requieren herramientas para la recuperación masiva de información, procesamiento analítico, extracción del conocimiento y visualización de los resultados.

En este trabajo se presenta un programa de computación para realizar análisis estandarizado de áreas de investigación y conceptos claves de la ciencia y la tecnología nucleares desarrollado en CUBAENERGÍA [1]. Se describen los componentes principales del sistema de procesamiento de información, así como los métodos computacionales y módulos para el análisis del contenido temático de archivos de registros de la base de datos INIS. Se exponen los resultados del análisis del contenido temático de la revista Nucleus desde su fundación hasta el presente. Además se muestran los resultados del estudio demostrativo Nucleus, Ciencia, Tecnología y Sociedad. El estudio abarcó todo el

periodo de existencia de la base de datos INIS, desde 1970 hasta la actualización más reciente disponible, septiembre 2006. En total se analizaron 55 992 registros documentales en formato de texto plano, 12,3 megabytes. Estos resultados aportan nuevos elementos para valorar el significado del contenido de la revista Nucleus en el contexto de la innovación en áreas de investigación multidisciplinarias interrelacionadas.

2. SISTEMA DE ANÁLISIS DE INFORMACIÓN

La figura 1 muestra el esquema de los principales componentes del sistema utilizado. La izquierda de la figura representa la aplicación desarrollada y suministrada por la sección INIS y Gestión del Conocimiento Nuclear del OIEA, instalada en el Servidor INIS Cuba. La base de datos bibliográficos INIS, la cual representa el estado del arte en el conocimiento nuclear documentado a nivel mundial, se tomó como la fuente estandarizada de información para el trabajo analítico [2]. La aplicación INIS Webspurs permitió recuperar la información sobre las áreas temáticas de interés y almacenarla en archivos de registros INIS. Los componentes de la derecha representan el programa para la recuperación de la información y el procesamiento analítico.

El interés se enfocó en el análisis del contenido temático de grandes colecciones de registros de documentos. El elemento clave en este tipo de análisis fue el campo de los descriptores. Cada término del campo de descriptores se consideró como una entidad observable e irreducible de contenido temático. El INIS Thesaurus se adoptó como definición del sistema de referencia terminológico analítico de cerca de 20 100 dimensiones [3]. El modelo del espacio vectorial de la matriz término-documento se utilizó para representar los archivos de registros INIS [4,5]. Para el procesamiento analítico y la visualización de la información se aplicaron los métodos computacionales del álgebra lineal y la geometría proyectiva.

2.1 Análisis temático unitario

Este módulo realizó el análisis de contenido básico de cada archivos de registros INIS como

unidad independiente. Calculó la proyección de todo el archivo de registros INIS en el INIS Thesaurus y permitió establecer el ordenamiento cuantitativo de cada descriptor temático de acuerdo con la medida de la proyección.

2.2 Análisis temático relacional

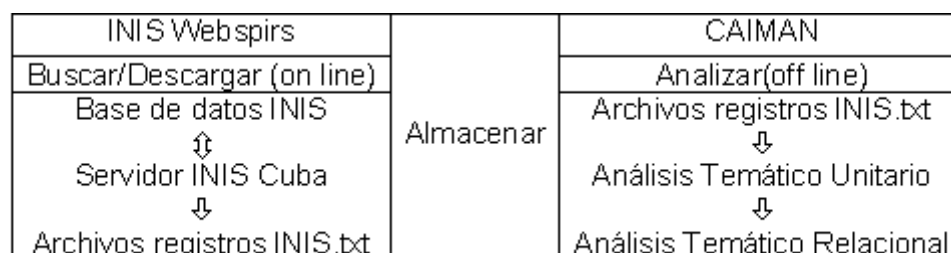
Este módulo realizó el análisis de la interrelación del contenido de todos los pares de áreas diferentes posibles. Para cada par de áreas efectuó la proyección diádica término-término en el INIS Thesaurus. Así determinó la medida del contenido temático común para cada par de áreas temáticas. Los cálculos proporcionaron la matriz relacional temática y sus direcciones propias proyectivas para el procesamiento analítico y la visualización.

3. RESULTADOS

3.1. Análisis temático unitario

En la tabla 1 están resumidos las áreas de investigación y los resultados del análisis temático unitario del estudio demostrativo Nucleus, Ciencia, Tecnología y Sociedad. Las áreas de investigación están caracterizadas en un sentido global por dos medidas, registros y términos. El número de registros, está asociado con la producción documental y sirve de base para la investigación bibliométrica. El número de términos es una medida del contenido temático y permite ganar comprensión acerca del conocimiento científico documentado. Como se reporta en la tabla 1, la revista Nucleus tiene 244 registros en la base de datos INIS. El contenido temático de todos los registros de Nucleus está descrito por 692 términos del INIS Thesaurus. En la propia tabla también están incluidos los resultados de otras áreas temáticas que representan conceptos claves de la ciencia, la tecnología y la sociedad. Estos resultados han permitido comparar las medidas de la revista Nucleus con el resto de las áreas estudiadas.

El espectro temático de Nucleus está mostrado en la figura 2. El número de orden y la frecuencia de ocurrencia de los términos están representados en los ejes horizontal y vertical, respectivamente. Ambos ejes tienen escalas



CAIMAN (Content Aimend Massive Analysis).

CIENCIAS NUCLEARES

Tabla 1. Análisis Temático Unitario. Estudio de Nucleus, Ciencia, Tecnología y Sociedad

Nº	Área	Registros	Términos
1	Nucleus	244	692
2	biology	1196	2844
3	chemistry	1869	3661
4	economy	1614	1972
5	education	8896	4330
6	engineering	3130	3255
7	environment	21667	7699
8	food	4818	3954
9	mathematics	2892	3620
10	public health	5747	3727
11	renewable energy sources	2852	1784
12	sustainable development	1067	1359

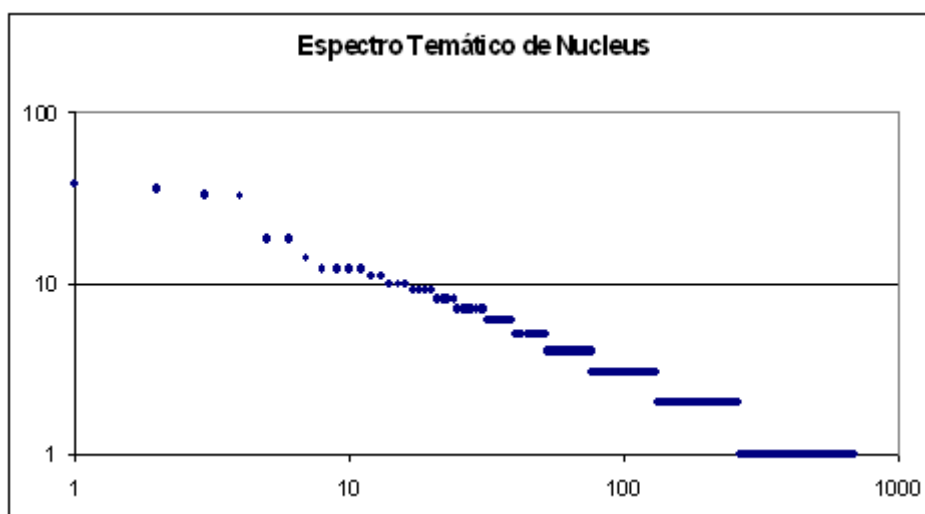


Figura 2. Espectro del contenido temático de la revista Nucleus.

logarítmicas. Se aprecia que los 10 primeros términos tienen frecuencias relativamente mayores que los restantes, es decir, son términos predominantes.

Período de 1986 al 2005

En la tabla 2, análisis top 10, aparecen reportados los 10 primeros términos y las frecuencias correspondientes a todo el período de existencia de Nucleus. El valor promedio de las frecuencias top 10, tabla 2, resultó igual a 22,4. La frecuencia en todo el espectro temático promedió el valor 2,1. La razón de predominancia de frecuencia fue igual a 10,5.

Paralización de la CEN Juraguá

En la tabla 3 está descrito el cambio del contenido temático de Nucleus a principios de la década del 90 cuando tuvo lugar la paralización de la

Tabla 2. Análisis Top 10. Todo el periodo de existencia de la revista de 1986 a 2005

Nº	Términos	F
1	radiation-doses	38
2	gamma-radiation	35
3	cuba	33
4	radiation-protection	32
5	cobalt-60	18
6	dose-response-relationships	18
7	quality-control	14
8	cesium-137	12
9	computer-codes	12
10	gamma-spectroscopy	12
Promedio F top 10		22,4
Promedio F espectro		2,1
Razón predominancia F		10,5

CIENCIAS NUCLEARES

Tabla 3. Análisis cambio temático. Periodos anterior y posterior a la paralización de construcción de la CEN Juaraguá

Análisis Temático Unitario				
Nº	Nucleus – Periodo	Años	Registros	Términos
1	1986 – 1992	7	74	256
2	1993 – 2005	13	170	532
Análisis Top 10				
Nº	Años 1986-1992	F	Años 1993 – 2005	F
1	cobalt-60	11	radiation-doses	37
2	gamma-radiation	11	radiation-protection	32
3	cuba	10	gamma-radiation	24
4	dose-response-relationships	7	cuba	23
5	computer-codes	6	quality-control	12
6	cesium-137	4	agriculture	11
7	neutron-activation-analysis	4	dose-response-relationships	11
8	structural-chemical-analysis	4	gamma-spectroscopy	11
9	cross-sections	3	plant-breeding	11
10	nuclear-magnetic-resonance	3	nuclear-medicine	10
Promedio F top 10		6,3		18,2
Promedio F espectro		1,4		2,1
Razón predominancia F		4,5		8,7

construcción de la central nuclear de Juaraguá. El cambio del contenido temático se evidencia a través del análisis temático unitario y del análisis top 10. Este análisis tiene únicamente carácter descriptivo. No trata de establecer relación causa efecto alguna entre la paralización de la construcción de la CEN de Juaraguá y cambio del contenido temático de Nucleus. Los estudios de estos cambios en el tiempo requieren aplicar métodos de análisis de series cronológicas.

3.2. Análisis temático relacional

La matriz relacional temática del estudio demostrativo Nucleus, Ciencia, Tecnología y Sociedad se muestra en la tabla 4. Las áreas temáticas analizadas están representadas por las filas y las columnas. Cada elemento de la matriz está determinado por el número de términos comunes al par de áreas asociadas con la fila y la columna. Las relaciones del contenido temático entre todos los pares de áreas están completamente descritas por la matriz. Así por ejemplo, el elemento en la fila 2, columna 1, significa que el par *biology, nucleus*, tiene 467 términos del INIS Thesaurus en común. Las medidas de las relaciones del contenido temático de la revista Nucleus con las áreas temáticas representativas de la ciencia, tecnología y sociedad están dadas por los elementos de la primera columna. Debido a la simetría de la relación entre las áreas temáticas, la matriz relacional es simétrica respecto a la diagonal y para mayor sencillez se omitieron todos los elementos superiores a la diagonal.

Por la dimensión del problema resulta difícil interpretar los elementos de la matriz de la tabla 4. Para facilitar la comprensión de las relaciones existentes entre las áreas temáticas se desarrolló un módulo de programa para transformar y visualizar la información contenida en la matriz relacional temática, basado en métodos computacionales del álgebra lineal y la geometría proyectiva. En la figura 3 está representado el mapa temático correspondiente a la matriz relacional de la tabla 4. Los puntos del mapa están numerados de acuerdo con las filas de la matriz. La configuración geométrica de los puntos del gráfico está determinada por la métrica del espacio vectorial lineal de la matriz relacional temática. De hecho, los puntos de la figura 2 son la mejor aproximación proyectiva bidimensional de la matriz de la tabla 4.

DISCUSIÓN

Los resultados del análisis temático unitario caracterizan satisfactoriamente el contenido temático de la revista Nucleus. Se identificaron en total 692 términos del INIS Thesaurus que describen completamente el total de 244 registros de Nucleus en la base de datos INIS.

El espectro temático, figura 2, evidencia que existen poco más de una decena de términos que son fuertemente predominantes. La tabla 2, análisis top 10, identifica los 10 primeros términos descriptores del contenido temático de Nucleus. El conjunto de estos 10 términos tiene una razón de predominancia de frecuencia de

CIENCIAS NUCLEARES

Tabla 4. Análisis temático relacional. Estudio de Nucleus, ciencia, tecnología y sociedad.

Nº	Área	1	2	3	4	5	6
1	nucleus	692					
2	biology	467	2844				
3	chemistry	462	1774	3661			
4	economy	309	890	1105	1972		
5	education	537	1817	2048	1423	4330	
6	engineering	413	1397	1822	1266	2147	3255
7	environment	623	2458	2912	1793	3530	2690
8	food	514	1786	1823	1216	2133	1521
9	mathematics	441	1527	1857	895	1960	1626
10	public health	496	1615	1754	1365	2417	1776
11	renewable energy sources	225	742	867	1085	1198	1051
12	sustainable development	227	649	729	897	978	831
Nº	Área	7	8	9	10	11	12
7	environment	7699					
8	food	3339	3954				
9	mathematics	2491	1511	3620			
10	public health	3357	2360	1428	3727		
11	renewable energy sources	1549	1059	631	1231	1784	
12	sustainable development	1240	914	518	1060	921	1359

10,7 sobre todo el espectro temático de 692 términos. Estos términos predominantes proporcionan la mejor descripción terminológica del contenido temático de Nucleus, usando 10 palabras del INIS Thesaurus.

La tabla 3 evidencia el cambio del contenido temático de Nucleus a principios de la década del 90, justo al final del periodo de 1986 a 1992, de siete años; y al inicio del periodo 1993 al 2005, de

13 años. Existió un apreciable crecimiento del contenido temático, el cual aumentó de 256 a 532, es decir, se extendió en un factor 2. La frecuencia promedio de todo el espectro temático aumentó de 1,4 a 2,1, es decir, creció en un factor de 1,5. El promedio de las frecuencias top 10 aumentó de 6,3 a 18,5, es decir, en un factor de 2,9. La razón de predominancia aumentó de 4,5 a 8,7, o sea, creció en un factor de 1,9. El aumento de la razón de predominancia indica la

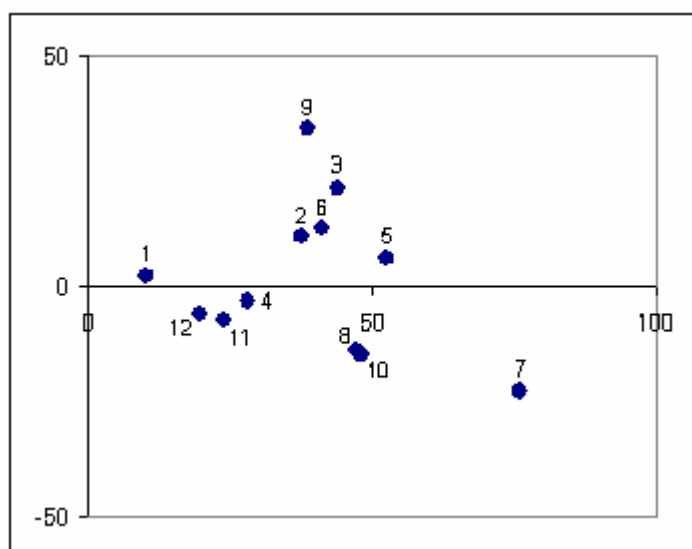


Figura 3. Mapa temático de Nucleus Ciencia, Tecnología y Sociedad. Numeración de las áreas temáticas, ver tabla 1

intensificación del contenido temático de los términos top 10. El cambio temático de Nucleus significó una duplicación tanto en la extensión de todo el espectro temático como en la intensidad de los términos top 10.

La matriz relacional temática del estudio demostrativo, tabla 4, muestra que el contenido de Nucleus está relacionado con los conceptos claves de la ciencia, tecnología y sociedad. La medida de la relación está dada por los elementos de la primera columna de la matriz. El valor máximo, igual a 623, corresponde al par (*nucleus, environment*). El valor mínimo, igual a 225, corresponde al par (*nucleus, renewable energy sources*). Si se toma el contenido temático de Nucleus como referencia, entonces las medidas relativas son (*nucleus, environment*), 90% y (*nucleus, renewable energy sources*), 33%. El 33% y el 90% del contenido temático de la revista Nucleus está relacionado con los conceptos claves considerados.

CONCLUSIONES

Se desarrolló un programa de computación capaz de realizar análisis estandarizado de áreas de investigación y conceptos claves de la ciencia y la tecnología nucleares a partir de archivos de registros de la base de datos INIS. Se analizó el contenido temático de la revista Nucleus. Se

identificaron en total 692 términos del INIS Thesaurus que describen completamente el total de 244 registros de Nucleus en la base de datos INIS. El estudio temático relacional demostrativo Nucleus, Ciencia, Tecnología y Sociedad aportó nuevos elementos para valorar el significado del contenido de la revista en el contexto de la innovación en áreas de investigación multidisciplinares interrelacionadas.

REFERENCIAS

- [1] GUERRA VALDÉS, R., Standardized analysis of research areas and key concepts of nuclear science and technology at CUBAENERGIA. IAEA-CN-153/3/P06, p. 136, Book of Extended Synopses, International Conference on Knowledge Management in Nuclear Facilities, 18-21 June 2007, Vienna, Austria.
- [2] International Atomic Energy Agency, INIS: Manual for Online Retrieval from INIS Database on the Internet, IAEA-INIS-24, IAEA, Vienna (2000).
- [3] International Atomic Energy Agency, INIS: Thesaurus, IAEA-INIS-13 (Rev.38), IAEA, Vienna (1999).
- [4] BERRY, M.W., DUMAIS, S.T., O2 BRIEN, G.W., Using linear algebra for intelligent information retrieval, SIAM Review, 37(4): (1995)573-595.
- [5] BERRY, M.W., DRMAË, Z., JESSUP, E. R., Matrices, vector spaces and information retrieval, SIAM Review, 41(2):(1999)335-362.

Recibido: 16 de octubre de 2007
Aceptado: 1 de noviembre de 2007