

Los terminos de indización y la investigación documental: Exposición de una aplicación concreta

Adelaida Román Román
CINDOC. CSIC

0.1. Resumen

Tras una consideración inicial sobre las posibilidades de los términos de indización como indicadores del conocimiento científico, se presenta un acercamiento al análisis y evolución de los temas estudiados por los científicos sociales españoles durante el periodo de la Transición Política (1976-1985), mediante el estudio estadístico de las ocurrencias de los términos de indización en la base de datos bibliográfica del ISOC.

Palabras clave: Análisis informetricos mediante descriptores. Lenguajes documentales.

0.2. Abstract

After a general consideration about indexing terms as indicators of scientific knowledge, this paper presents an approach to the analysis and evolution of subjects studied by social scientists in Spain during Political Transition (1976-1985), based on occurrences of indexing terms in the ISOC bibliographical data base.

Keywords: Indexing terms for informetrics. Indexing languages.

1. Introducción

Resuelto el problema previo de la relación de pertinencia entre los términos de indización y los contenidos objetivos de los textos científicos, esta comunicación pretende subrayar, más allá de la función principal de los términos de indización, que es la de representar el contenido conceptual de los documentos, y a partir de ella, el interés que tienen éstos para la investigación documental. Constituyen los elementos básicos para detectar las orientaciones de la investigación en cualquier campo del conocimiento y para seguir su evolución.

Así, podríamos considerar con Xavier Polanco, que los términos de indización son verdaderos “indicadores del conocimiento”.

En efecto, las Base de Datos nos proporcionan el acceso a grandes cantidades de información en las que los contenidos científicos aparecen, de alguna manera, yuxtapuestos. Pero el conocimiento científico no está hecho de yuxtaposición de elementos, sino más bien de redes de asociaciones múltiples entre conceptos, técnicas, instrumentos, métodos, campos de aplicación, etc., y los términos de indicación, en un ámbito temático dado, pueden ponernos de manifiesto el entramado de relaciones en ese campo científico. Pueden permitirnos estudiar la evolución en el tiempo de las investigaciones sobre un tema, el surgimiento del interés por temas nuevos, las relaciones entre enfoques diferentes que se estudian por separado, la evolución en las técnicas de análisis o de experimentación, etc.

Estas posibilidades abren perspectivas interesantes de estudio aplicables bien a la investigación bien a la toma de decisiones no sólo en el campo de la política científica sino a las decisiones estratégicas en el ámbito empresarial.

2. Exposición de una aplicación concreta: Análisis y evolución de los temas estudiados por los científicos sociales en España durante la Transición Política (1976-1985), a partir de los términos de indicación.

En el marco de un estudio más amplio que tiene por objeto establecer en qué medida la producción bibliográfica de los científicos sociales es representativa de los problemas y preocupaciones de la sociedad de su época, se intenta, como primera etapa del análisis, determinar cuáles son los intereses de los autores, expresados en sus publicaciones y sintetizados en los descriptores que representan los contenidos conceptuales de esas publicaciones, indizadas en la B. D. ISOC.

Se analizan también las variaciones de esos intereses a lo largo del período estudiado.

Se parte de una muestra cuyas características son las siguientes:

- Número total de documentos.....11.045 art. de rev.
- Número de descriptores utilizados (76-80).....22.519
- Número de descriptores utilizados (81-85).....48.189
- Número total de descriptores diferentes.....15.725
- Media desc./documento (76-80).....6,25
- Media desc./documento (81-85).....6,47

Los artículos fueron extraídos de 32 títulos de revistas previamente seleccionadas de manera que todos los campos relativos a las CC. Sociales estuvieran representados. Se dividió el corpus documental en dos períodos cronológicos iguales con el objeto de poder observar las dinámicas y los cambios de interés de los autores.

En primer lugar se analizan los contenidos generales dominantes en cada uno de los dos períodos y en el conjunto de años estudiados. Todo el trabajo se realiza partiendo de la hipótesis de que el interés o la importancia de un tema cualquiera está en relación directa con su *frecuencia de aparición* en el corpus que se estudia. De ahí se deduce que el interés de un tema en este caso viene expresado por la frecuencia de aparición de los descriptores que lo representan.

Dado que la frecuencia de los términos sigue la Ley de Zipf, según la cual la frecuencia es inversamente proporcional al número de términos que tienen una frecuencia dada, (es decir, a mayor frecuencia, menor número de términos con esa frecuencia), se procedió a estructurar los términos de indización en rangos de frecuencia. Se definieron ocho rangos o tramos:

- Tramo 1: descriptores que aparecen más de 100 veces
- Tramo 2: descriptores que aparecen entre 75 y 100 veces
- Tramo 3: descriptores que aparecen entre 74 y 51 veces
- Tramo 4: descriptores que aparecen entre 50 y 41 veces
- Tramo 5: descriptores que aparecen entre 40 y 31 veces
- Tramo 6: descriptores que aparecen entre 30 y 21 veces
- Tramo 7: descriptores que aparecen entre 20 y 11 veces
- Tramo 8: descriptores que aparecen menos de 11 veces

Por otra parte se procedió a categorizar los 15.725 descriptores según su pertenencia a uno de cuatro grandes ámbitos temáticos de carácter muy general: ámbito económico, socio-cultural, jurídico-político y neutro.

Para establecer comparaciones entre los diferentes rangos de frecuencia para los dos períodos, en relación con el peso que cada temática general tenía en cada uno de ellos, y poder expresarlas cuantitativamente, se procedió a asignar un peso a cada término de indización. Se define el *peso* del descriptor A como el valor de su frecuencia dividido por la suma de las frecuencias de los demás términos de su tramo y multiplicado por cien. El resultado se expresa, pues, porcentualmente.

Los resultados obtenidos una vez calculado el peso de cada grupo temático general en cada tramo, para cada uno de los dos períodos estudiados pueden verse en los gráficos 1, 2 y 3. En ellos puede observarse que:

1º) Los descriptores relativos al ámbito económico tienen un peso muy importante y sostenido a lo largo de todo el período.

2º) Los descriptores relativos al ámbito socio-cultural aumentan de una manera clara en el segundo período.

3º) Los descriptores del ámbito jurídico político, ausentes en los primeros años, hacen sentir su presencia en los cinco últimos.

Hay que tener en cuenta en la interpretación de los gráficos, que el peso para cada período lo marcan especialmente los primeros tramos que son los de frecuencias más altas. Al disminuir las frecuencias, aumentan los términos y también la presencia de temas es más variada, pero esta presencia es menos significativa.

3. Dinámica de la información

Para estudiar la dinámica de la información en el período acotado, se analizan las migraciones de los términos siempre a partir del segundo período, para detectar, en cada tramo, cual es el tramo de origen de cada uno de los términos en el primer período. Este procedimiento puede verse en los gráficos 4 y 5 que expresan esta dinámica para los dos primeros tramos.

A partir de este análisis se constata que se produjeron cambios significativos en el interés de los autores con el transcurso de los años: así el proceso de integración de España en la CEE, la reforma fiscal, la crisis económica, el desempleo, la política económica o el comercio exterior son temas cuyo interés se sitúa en primera fila a lo largo de todo el período. Pero además se constatan cambios: hay temas que al principio, en los 70, apenas sí interesan y que afloran con fuerza en los ochenta. Y, al contrario, intereses muy presentes al principio, llegan a casi desaparecer en los últimos años: la enseñanza, la juventud, el cambio tecnológico, y, en general, los temas socio culturales aparecen con fuerza hacia 1980 y van a tener una presencia cada vez mayor según avanza el período. Surgen también temas nuevos, que apenas se detectaban en las publicaciones de 1976 a 1980: marginación social, drogadicción, derechos sindicales, educación compensatoria, fracaso escolar... Otros temas con frecuencias muy altas en los primeros años, han casi desaparecido de la literatura publicada con posterioridad a 1981: Desarrollo económico, política agraria, ideología política, marxismo ... son algunos ejemplos.

4. Conclusión

Podemos afirmar que el método del análisis de frecuencias aplicado a los términos de indización de los documentos, y la distribución de éstos en rangos, han permitido estudiar las migraciones de los términos a lo largo del período acotado, facilitando el análisis de la evolución de los intereses de los autores y ofreciendo una posibilidad de cuantificación de sus preferencias manifestadas en la literatura científica a lo largo del período. La aplicación presentada es sólo un ejemplo, que sirve para ilustrar algunas de las posibilidades que ofrecen los términos de indización para el estudio y seguimiento de las tendencias de la investigación en cualquier campo científico o tecnológico.

5. Bibliografía

Polanco, Xavier ; Grivel, Luc ; François, Claire; Besagni, Dominique. L'informétrie, un programme de recherche. // Les systèmes d'information élaborée : Journées SFBA, 1993, 9-11 Juin, Ile Rousse.

Braga, G. A. A model for evaluating scientific and technical journals from developing countries. New York : Knowledge Industry Publications Inc. for the American Society of Information Science, 1981.

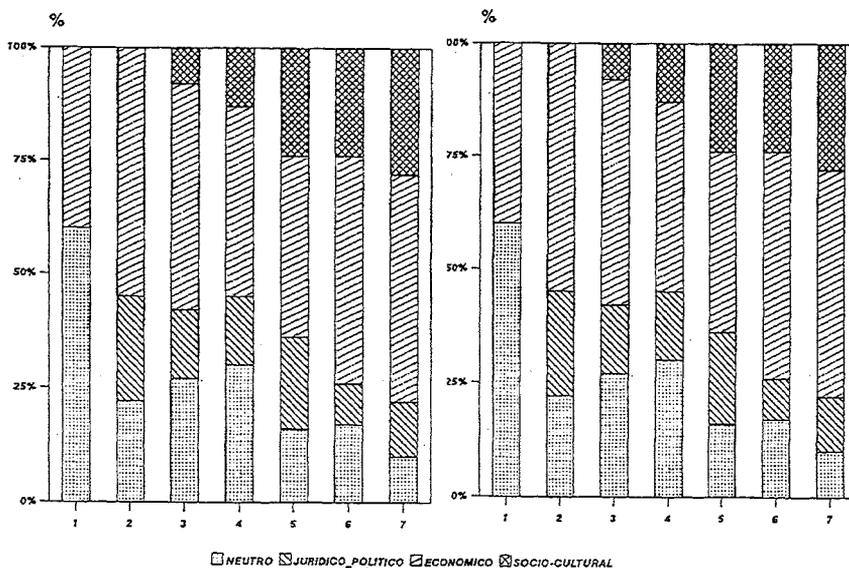
Dutheuil, C.. Stratégie d' interrogation et interpretation des resultats de l'analyse des données, ou de la Banque de Données à la bibliométrie. // Journées ADBS, 1898, Mayo 11, Orsay.

Dutheuil, C. Les problèmes de l'hétérogénéité des informations à soumettre à l'analyse des données. // Journées de la Société Française de Bibliométrie Appliquée : 1987, Septembre, Ile Rousse.

Ohly, H. Peter. A procedure for comparing documentation language application : the transformed Zipf curve. // International Classification. 9 : 3 (1982) 125-128.

Oppenheim, C. Use of on-line databases in bibliometric studies. // On Line Information Meeting, 1985, december, London.

6. Gráficos



Evolución de los temas de interés en el primer y segundo quinquenios

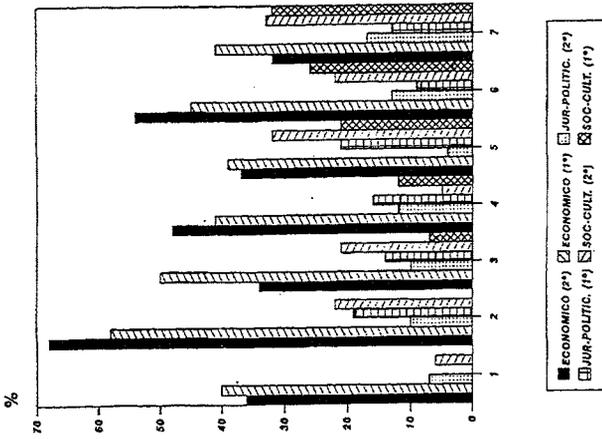
FLUJOS DE INFORMACION: ORIGEN DE LOS TERMINOS QUE FORMAN EL TRAMO 2 EN EL 2º QUINQUENIO.

| TRAMOS | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------------------|------------------------------------|--------------|--|-------------|-------------|-------------|-------------|
| FRECUENCIAS | X > 100 | 100 > X > 75 | 75 > X > 50 | 50 > X > 40 | 40 > X > 30 | 30 > X > 20 | 20 > X > 10 |
| Nº DECS. 1º QUINQUENIO | 2 | 1 | 4 | 0 | 1 | 0 | 1 |
| | HABIENDA PUBLICA SISTEMA FISCAL | LEGISLACION | DESEMPEÑO POLITICA FISCAL POLITICA MONETARIA PIRE | JUVENTUD | | | ENSEÑANZA |
| Nº DECS. 2º QUINQUENIO | | 10 | | | | | |

NUEVO TERMINO: EXPERIENCIA PEDAGOGICA

FLUJOS DE INFORMACION: ORIGEN DE LOS TERMINOS QUE FORMAN EL TRAMO 1 EN EL 2º QUINQUENIO.

| TRAMOS | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------------------|---|------------------|---------------------------------------|-------------|-------------|-------------|--------------------|
| FRECUENCIAS | X > 100 | 100 > X > 75 | 75 > X > 50 | 50 > X > 40 | 40 > X > 30 | 30 > X > 20 | 20 > X > 10 |
| Nº DECS. 1º QUINQUENIO | 3 | 1 | 2 | 0 | 0 | 0 | 1 |
| | CEE POLITICA ECONOMICA TEORIA ECONOMICA | CRISIS ECONOMICA | COMUNIDADES AUTONOMAS FINANCIACION | | | | ENSEÑANZA PRIMARIA |
| Nº DECS. 2º QUINQUENIO | | | | | | | 7 |



Evolución de los temas de interés: comparación de quinquenios