

COMUNICACIONES LIBRES

LAS BASES DE DATOS COMO FUENTES DE INFORMACIÓN PARA ESTUDIOS BIBLIOMÉTRICOS

GALBAN, C.

VAZQUEZ, M.

Instituto de Información y Documentación
en Ciencia y Tecnología. (ICYT). CSIC. Madrid

RESUMEN

Las bases de datos constituyen, en principio, una fuente de información excepcional para la realización de estudios bibliométricos.

Se expone la metodología seguida por los autores para el manejo de las referencias obtenidas, y los problemas que se derivan de la falta de homogeneidad y de normalización de la información contenida en los registros de las bases de datos que se utilizan en los citados estudios.

Se presentan ejemplos que ilustran estas dificultades, tomados de estudios de la producción científica de diversos colectivos españoles.

LAS BASES DE DATOS COMO FUENTES DE INFORMACIÓN PARA ESTUDIOS BIBLIOMÉTRICOS

Las bases de datos bibliográficas, tanto en su versión impresa, la revista de resúmenes, como en su forma automatizada, se utilizan principalmente para recuperar referencias específicas relativas a la información sobre un *tema determinado*, publicada en diversos tipos de documentos primarios. También se emplean con frecuencia para recuperar las referencias correspondientes a un autor dado.

Sin embargo estas bases de datos contienen mucha más información en cada referencia: país e institución en que se ha realizado el trabajo, idioma de publicación, tipo de documento a que se refiere, etc., que puede ser recuperada cuando se consultan de forma automatizada, lo que hace posible su empleo en aplicaciones tales como estudiar la productividad y hábito de publicación de un grupo o equipo científico, determinar las revistas que acogen los trabajos relativos a un tema específico, hallar el grado de cobertura de una publicación por varias bases de datos, etc.

Esto es debido a que los registros de la mayoría de las bases de datos contienen una serie de campos, comunes a todas ellas, que incluyan información recuperable de forma automatizada. Estos campos son:

- Título
- Autor (es)
- Institución o afiliación de los autores (incluyendo el país)
- Fuente (incluyendo título abreviado o completo, CODEN o ISSN)
- Año de publicación
- Descriptores de indización

Ejemplo 1

Section 24 (ANESTHESIOLOGY)
 Volume 15 Issue 8 Abstract Number

DIALOG Accession Number 80054721
 EM Abstract Number 0240150802210

Acute intoxication with Verapamil }
 INTOXICATION AIGUE PAR LE VERAPAMIL } /TI

AU= Haegy J.M.; Hasselmann M.; Faller J.P.; et al.
 CS= { Serv. Reanimat. Med., CHU, Strasbourg } CP-
 JN= J. MED. STRASBOURG (FRANCE), 1979, 10/3-4 (157-160), PY-
 CO= Coden: JMSTB

LA= Languages: FRENCH Summary Languages: ENGLISH SL-

The text describes a case of acute intoxication due to 1.2 g of Verapamil. This produced severe symptoms of auriculo-ventricular conduction defect. These were reversible after three days. The patient was a young girl of 14 years. The authors analyse the main aspects of acute intoxication with this antiarrhythmic calcium inhibitor drug. In the healthy adult a dose of 0.6 to 1g of Verapamil can be considered to be toxic. The symptomatology is dominated by auriculo-ventricular conduction disorders and the risk of cardiac failure. The treatment is symptomatic and direct. The role of drugs such as atropine, beta-stimulants digitalis, calcium salts, glucagon and electrical pacing are discussed. /AI

MN= Manufact Names: fra/knoll; roche /TI
 TC= Trade Names: isoptine; mogadon; isuprel /CS
 Tags: CASE REPORT(S)(0151); HEART(0921); ORAL(0181); DRUG
 ADVERSE REACTIONS(0198); INTOXICATION (POISONING)(0302)

DC= Descriptors: *verapamil(0051463); *heart atrioventricular block(0076778); *drug overdose(0195019); *adverse drug reaction(0055863); *intoxication(0024335); *nitrazepam(0033247); *atropine(0004087); *isoprenaline(0024906); *dopamine(0013981); *ouabain(0034892); *ajmaline(0001334); *disopyramide(0013819) /II
 Identifiers: girl of 14; case; /ID

SH= Section Headings: 0382100000-ADVERSE REACTIONS TITLES/DRUGS ACTING ON THE HEART /SH
 0371001000-DRUGDOC/DRUGS AFFECTING THE CARDIOVASCULAR SYSTEM/Antiarrhythmic and arrhythmia inducing drugs
 03726060100-/TOXIC SUBSTANCES AND PROTECTING AGENTS/Drugs/Toxic

Además de estos campos comunes, existen otros específicos de la temática o especialidad de alguna base de datos determinada (por ejemplo, clasificación sistemática o taxonómica, etc.).

En el Instituto de Información y Documentación en Ciencia y Tecnología se han realizado diversos estudios bibliométricos y sociométricos aprovechando estas posibilidades que ofrecen las bases de datos bibliográficos (1, 2). Ahora bien, aun cuando con estos métodos se pueden recuperar rápidamente referencias cuya recogida manual sería laboriosísima e incluso a veces imposible, por la limitación de los fondos bibliográficos fácilmente disponibles y la falta de vías de entrada a los distintos aspectos a estudiar, presentan una serie de dificultades originadas por la deficiente o escasa normalización y uniformidad de los datos y registros de esas bases de datos.

Estas dificultades hacen que en muchos casos las conclusiones sacadas de estos trabajos estén fuertemente condicionadas por la metodología empleada, que a su vez depende de las características de cada base de datos.

Así por ejemplo, cuando un trabajo ha sido realizado por autores pertenecientes a dos o más instituciones, sólo aparece normalmente la del autor que figura en primer lugar. También puede ocurrir que no aparezca el país en que se ha realizado aunque sí la ciudad, o que la asignación del país se haga de una forma errónea.

Ejemplo 2

TYPE 5/5/7
84-X-0308024 PASCAL 84606735
PERSPECTIVA LOGICO-LINGUEISTICA DE LA DOCUMENTACION
- PERSPECTIVE LOGICO-LINGUISTIQUE DE LA DOCUMENTATION -
GUITARD (A. L.)
AFF: C.S.I.C., INST. INFORMACION DOCUMENTACION CIENCIA TECNOLOGIA
(ICYT)/MADRID/USA ←
REV. ESP. DOC. CIENT.//REVISTA ESPANOLA DE DOCUMENTACION
CIENTIFICA; 0210-0614; ESP; Date: 1984; Vol: 7; No: 1; p.: 45-51; 10
REF.; Cote: 17953; Langue: Espagnol Type: TP, L
CC: 205.A.01.A

TYPE 5/5/80
79-5-0021343 BIOPASCAL 79012096
ESTUDIO SOBRE LA ACCION TERAPEUTICA DE LA VASOPRESINA EN SINDROMES
AMNESICOS Y AVANCE SOBRE SU POSIBLE ACCION EN LAS DEPRESIONES.
- ETUDE SUR L'ACTION THERAPEUTIQUE DE LA VASOPRESSINE DANS LES
SYNDROMES AMNESIQUES ETHYPOTHESE SUR SON ACTION POSSIBL DANS LES
DEPRESSIONS. -
OLIVEROS (J. C.); JANDALI (M. K.)
AFF: HOSP. CLIN. MADRID ←
ARCH. NEUROBIOL.; ESP; Date: 1978; Vol: 41; No: 1; p.: 63-74;
Resume: ANGL.; 16 REF.; Cote: 8175; Langue: Espagnol Type: TP, LA
CC: 330.A.02.B.03.C
? 13/78

Es preciso indicar aquí el notable esfuerzo que a este efecto están realizando los productores y distribuidores de bases de datos en los últimos años, para subsanar estos errores que presentan. Sin embargo, todavía hoy estamos lejos de decir que estén superadas todas las dificultades mencionadas.

En esta comunicación vamos a explicar la sistemática utilizada en los trabajos bibliométricos realizados para estudiar la producción científica de una comunidad o equipo determinados. Se han utilizado para ello las referencias obtenidas en la consulta a bases de datos bibliográficos internacionales.

La recuperación suele hacerse por el campo de afiliación de los autores, que en la mayor parte de las bases de datos se designa como «corporate source».

Ejemplo 3

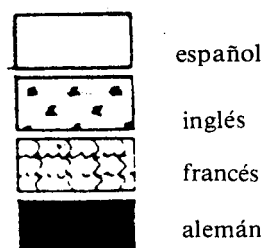
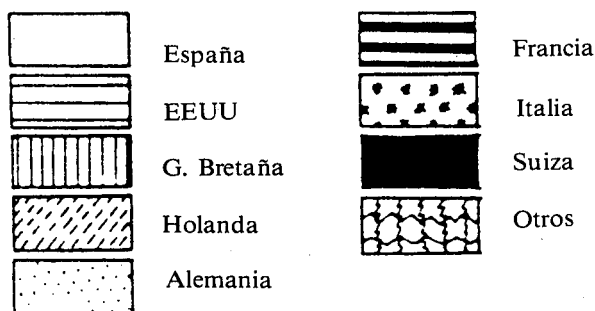
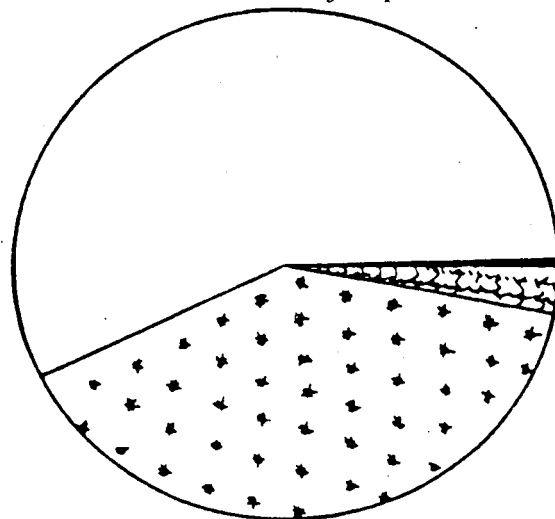
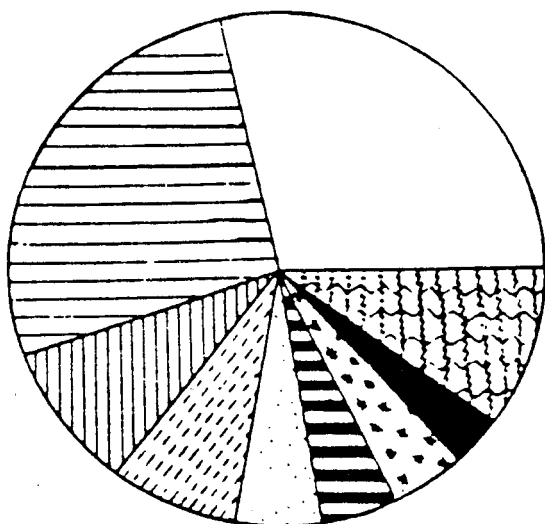
```
? s zarag?  
 1      10 ZARAG?  
? s lall/104-108  
      LIMIT ALL/104-108  
? s cs=zarag?  
 2      327 CS=ZARAG?  
? s cs=sarag?  
 3          3 CS=SARAG?  
? c2+3  
 4      328 2+3  
? pr4/t 4/5/1
```

Una vez obtenidas las referencias correspondientes a la institución, ciudad o país que se desea, es necesario trabajar con ellas usando la versión impresa, ya que la falta de homogeneidad de los datos, que se pone de manifiesto en los distintos ejemplos que aquí se dan, —tomados todos de diferentes estudios bibliométricos realizados en el ICYT—, hace que el tratamiento automático de dichas referencias dé lugar a numerosos errores. En los casos en que el distribuidor de las bases de datos ofrezca esa posibilidad, y con el fin de facilitar el manejo de las referencias, se solicitan éstas ordenadas alfabéticamente por la fuente de publicación. Esta es una clasificación que en un momento u otro habrá que llevar a cabo, por ser la más apropiada para eliminar referencias repetidas dentro de la misma base, en el caso de que las hubiera, y más tarde también para detectar posibles solapamientos entre diferentes bases de datos, se han consultado varias (3).

De las referencias así clasificadas se obtienen datos relativos a la distribución en función del tipo de documento (artículo de revista, comunicación a congreso, patente, etc.) e idioma de publicación y, para el caso de las revistas, por frecuencia de uso, país de edición, etc., que permiten determinar los hábitos de publicación de la comunidad científica en estudio.

Distribución de las revistas por países

Distribución de los trabajos por idiomas



Ejemplo 4. Distribución de los trabajos realizados en Salamanca en 1980-1983 en función del idioma de publicación y de las revistas utilizadas por país de origen.

Cuando el estudio se realiza sobre una comunidad muy amplia, que puede ser: país, ciudad, universidad, etc., se procede a clasificar las referencias en función de la institución en que se han realizado. Se distribuyen en primer lugar en grupos correspondientes a *instituciones de primer nivel*: Universidad, Organismos Públicos de Investigación (OPIs), Empresas Privadas, etc. Después se procede a clasificar cada uno de estos grupos en otros más específicos, considerados como *instituciones de segundo nivel*, por ejemplo las referencias correspondientes a la Universidad se distribuyen en facultades y, dentro de éstas, en departamentos (tercer nivel).

Ejemplo 5.

- a) Distribución de los trabajos de Zaragoza recogidos en el Chemical Abstracts en función de las Instituciones en que se han realizado.

Institución	N.º trabajos	%
Universidad	722	84,8
CSIC	121	14,2
Otros	3	0,4
Empresas Privadas	5	0,6
Total	851	100,0

- b) Distribución de los trabajos de Zaragoza realizados en centros del CSIC.

Centro	N.º trabajos	% CSIC	% total Zaragoza
Est. Exp. Aula Dei	13	1	1,5
Inst. Carboquímica	57	47	6,7
Inst. Ciencia Mater.	51	42	6,0
Total	121	100	14,2

Esta parte del trabajo presenta muchas dificultades, ya que estos datos no siempre figuran en las referencias, bien porque los propios autores o las revistas no los han indicado, o porque las bases de datos no los recogen (4). Cuando el dato aparece, puede figurar de muy diferentes maneras debido a la falta de normalización, por lo que es necesaria su homogeneización.

99: 90644j FCC process for catalytic cracking of gas oil to obtain gasolines. V. Design of a continuous fluidized bed pilot plant. Comparison of catalysts. Corella, J.; Iranzo, C.; Fernandez, Ana; Labordeta, L. (Univ. Zaragoza, Zaragoza, Spain). *Ing. Quim. (Madrid)* 1983, 15(170), 53-63 (Span). The activity, selectivity, and deactivation of various com. cracking catalysts were studied in a small pilot plant for catalytic cracking of up to 2 L/h gas oil to obtain gasoline. The system is based on a fluidized bed with continuous catalyst feed and without regeneration. Three cracking catalysts (MZ-6S [86836-40-0], MZ-7P [86836-39-7], and CBZ-1 [62362-86-1]) were studied at 500° cracking temp. under variable solids residence time, catalyst supply/gas oil ratio, and spatial velocity conditions. A microkinetic model of the reactor was used to simplify calcns. The catalysts in order of decreasing activity and resistance to deactivation were MZ-7P > MZ-6S > CBZ-1. In terms of selectivity, the most active catalyst in regard to gasoline formation was MZ-7P, followed by CBZ-1 and MZ-6S.

96: 180528s Initial rate of oxidation of isobutene over a mercuric chloride catalyst. Corella, J.; Asua, J. M. (Fac. Cienc., Univ. Zaragoza, Spain). *An. Quim., Ser. A* 1981, 77(1), 137-43 (Span). The kinetics and mechanism of the title autoxidn., to CH_2CMeCHO (I), was examd. at 180-200° in a fixed bed differential reactor contg. an active C supported HgCl_2 catalyst. This reaction also competes with the autoxidn. to AcCH:CH_2 (II) and the total autoxidn. (to CO_2 and H_2O). The activation energies for the autoxidn. decrease in the product order given: $\text{CO}_2\text{-H}_2\text{O} > \text{I} \gg \text{II}$.

104: 91237n Flow model for the solid in a continuous fluidized bed with increase of the cross section in its upper zone. Corella, Jose; Bilbao, Rafael; Monzon, Antonio; Lezaun, Javier; Fernandez, Fernando (Fac. Sci., Univ. Zaragoza, Zaragoza, Spain). *Ind. Eng. Chem. Process Des. Dev.* 1986, 25(1), 188-97 (Eng). A model of the flow of the solids in a continuous fluidized bed was developed and tested. Four different fluidized beds with a conical enlargement of the cross section in the upper zone of the bed were used. The model is a combined or compartment one and has 3 zones: a piston flow with dispersion, a backmixing flow in series, and a dead zone. The parameters of the model are calcd. from the residence-time-distribution curves obtained exptl. under various operating conditions. A small increase of the gas velocity at the bed surface increased the dispersion no. considerably. An increase of the solid fraction in the upper zone increased the solid circulating in piston flow. Internals in the tronconical zone decreased the dead zones and the dispersion no.

96: 206071q Composition and surface acidity of silicon dioxide-aluminum oxide catalysts prepared by impregnation and coprecipitation-impregnation. Corella Tudanca, Jose; Bilbao Elorriaga, Javier; Pilar Aznar, Maria (Dep. Quim. Tec., Univ. Zaragoza, Zaragoza, Spain). *Afinidad* 1981, 38(376), 518-21 (Eng). Several $\text{SiO}_2\text{-Al}_2\text{O}_3$ catalysts are prepd. by

impregnation and copptn.-impregnation. By using both methods to prep. the silica-alumina gel, the effects of the impregnating soln. concn., the support particle size, and the support aging time on the catalysts compn. and its surface acidity were studied

105: 117722q The modeling of the kinetics of deactivation of monofunctional catalysts with an acid strength distribution in their nonhomogeneous surface. Application to the deactivation of commercial catalysts in the FCC process. Corella, Jose; Menendez, Miguel (Dep. Chem. Eng., Univ. Zaragoza, Zaragoza, Spain 50009). *Chem. Eng. Sci.* 1986, 41(7), 1817-26 (Eng). A model for the kinetics of the deactivation of monofunctional catalysts with a heterogeneous or nonuniform active surface is presented. This model is based on previous works by Butt et al. (1978) and starts from the consideration of a function of strength distribution for the active sites. Four types of strength distribution are considered. In this model, time-on-stream is used as the independent variable. The influence of the values of the parameters of the catalyst surface on the activity-time curves and the parametric sensitivity of the model are analyzed. With the consideration of the nonuniformity of the catalyst surface developed in this model, one can adjust perfectly well the data of deactivation of com. catalysts in the cracking of gas oils and the empirical correlation of a general character proposed by Pacheco and Petersen (1984) for a large variety of processes and catalysts.

Ejemplo 6

Como estos estudios se han realizado casi siempre en el ICYT sobre personal investigador de los Centros públicos, elaborados por el Ministerio de Educación y Ciencia (6). Sin embargo, no todos los casos pueden resolverse satisfactoriamente.

Las dificultades mencionadas para la distribución por instituciones se multiplican cuando se quiere hacer un estudio de los autores, por ejemplo para determinar los más productivos, ya que pueden aparecer identificados de muy diferentes maneras, en particular cuando firma sus trabajos utilizando dos apellidos, caso muy frecuente en autores españoles. Así, hemos encontrado el caso de un investigador que en una misma base de datos aparece de nueve formas diferentes.

García Fierro J.L.
García Fierro, José Luis
García-Fierro, J.L.
Fierro, G.
Fierro, J.L.G.
Fierro, J.L. García
Fierro, José L.G.
Fierro, José Luis G.
Fierro, José Luis García.

Ejemplo 7

Cuando los dos apellidos son muy comunes aumentan las probabilidades de cometer errores, tanto por exceso como por defecto.

En general, las deficiencias de las bases de datos a este respecto, hacen que para estudios de productividad de autores que se quieran hacer con garantías de exhaustividad en la recuperación, sea necesaria la comparación con los datos tomados del curriculum de los propios autores.

Estos estudios no se realizan consultando a una sola base de datos, sino al mayor número de bases de datos bibliográficos posibles, que por su temática puedan incluir la información buscada. Por ejemplo, una consulta relativa al campo de la Farmacia deberá hacerse en: Pharmaceutical Abstracts, Index Medicus, Excerpta Medica, Biological Abstracts, Chemical Abstracts, etc. Aunque el trabajo se complica, la metodología es muy semejante a la anteriormente descrita. En primer lugar, se hace el estudio de los resultados obtenidos en cada base de datos por separado y después se comparan y eliminan las referencias recogidas en más de una base, a fin de obtener un conjunto de referencias únicas con el que se procede de la forma indicada anteriormente. Posteriormente se pueden hacer estudios de solapamiento de diferentes bases de datos, determinación del porcentaje de referencias que se recogerían consultando una, dos, tres, o más bases, y comparación de las bases consultadas en cuanto a cobertura temática, cobertura geográfica, período de tiempo necesario para la inclusión de referencia, etc.

Ejemplo 8

a) Solapamiento en varias bases de datos de los trabajos realizados en Salamanca en 1980-1983.

	BIOSIS	CA SEARCH	CAB	EMBASE	FSTA	GEOREF	INSPEC	I.P.A.	I.R.L.	MEDLINE	SCISEARCH
BIOSIS	260	77	33	87	4	3	0	16	25	43	122
CA SEARCH	77	346	46	74	7	14	13	17	25	37	151
CAB	33	46	108	7	4	2	5	6	22		
EMBASE	87	74	7	250	0	0	2	21	30	55	130
FSTA	4	7	4	0	11	0	0	0	0	1	6
GEOREF	3	14	2	0	0	35	0	0	0	0	6
INSPEC	3	13	0	2	0	0	32	0	0	0	18
I.P.A.	16	17	2	21	0	0	0	32	1	2	16
I.R.L.	25	25	5	30	0	0	0	1	58	15	37
MEDLINE	43	37	6	55	1	0	0	2	15	76	57
SCISEARCH	122	151	22	130	6	6	18	16	37	57	433

b) Recogida en las diversas bases de datos de los trabajos realizados en Salamanca en 1980-1983

Bases de datos	Referencias totales	Referencias exclusivas	% de ref. exclusivas de cada base	% del total de trabajos no repetidos (915)
BIOSIS	260	71	27,30	7,76
CHEM. ABS.	346	103	29,76	11,26
CAB	108	36	33,3	3,93
EMBASE	250	61	24,40	6,67
FSTA	11	4	36,36	0,44
GEOREF	35	16	45,71	1,75
INSPEC	32	10	31,25	1,09
Int. Pharm. Abs.	32	1	3,12	0,10
Life Sci. Coll	58	3	5,17	0,33
MEDLINE	76	3	3,95	0,33
SCISEARCH	433	141	32,56	15,41

Estos estudios se han hecho hasta ahora de forma manual. Es decir, las referencias obtenidas por consulta automatizada a las bases de datos, se tratan de forma manual una vez recibidos los listados fuera de línea para: determinación de solapamientos, recuentos de trabajos procedentes de las distintas instituciones, distribución por idiomas, etc. Sin embargo, los últimos desarrollos tecnológicos que han dado lugar a la puesta a punto de procedimientos de teledescarga, es decir, de transferencia de ficheros de un ordenador a otro, ofrecen una serie de posibilidades entre las que se encuentra el poder reunir en un archivo único las respuestas obtenidas a una consulta hecha en varias bases de datos, cuando para ello se ha empleado un terminal inteligente. Así es posible reestructurarlas o clasificarlas más tarde fuera de línea, a fin de obtener la información clasificada por los aspectos que interesen en particular. Con ello se eliminarían las operaciones manuales, tan tediosas y lentas (5).

Sin embargo, para que esto sea una realidad es necesario que además de superarse los problemas técnicos, legales y económicos que aún quedan por resolver, se lleve a cabo una profunda labor de normalización de registros y de unificación de criterios, por parte de los productores y los distribuidores de las bases de datos, los editores de las publicaciones primarias y los propios autores de los trabajos. En la situación actual de falta de homogeneidad y de normalización con que aparecen los citados campos en los registros bibliográficos, los valores que se obtienen con el tratamiento automático de los datos, aunque objetivos, no dan una visión real de los aspectos estudiados.

BIBLIOGRAFÍA

- (1) VAZQUEZ, Manuela; SANCHO, Rosa. «Estudio de la producción científica española sobre polímeros en el período 1974-1979». *Revista de plásticos modernos*, 40 (1980), p. 713-720.
- (2) LOPEZ AGUADO, Godofredo; DE LA VIESCA, Rosa. «Trabajos científicos publicados por los físicos españoles y recogidos en el INSPEC durante 1978-1979». *Revista española de documentación científica*, 6 (1983), p. 317-330.
- (3) GALBAN, Camen; VAZQUEZ, Manuela; DE LA VIESCA, Rosa; LAGUENS, Javier. «La producción científica de Salamanca 1980-1983». *Revista española de documentación científica*, 8 (1985), p. 321-348.
- (4) GOMEZ, Isabel; GALBAN, Carmen. «Lack of standardisation in the corporate source field of different databases». 10th On Line Meeting, 1986, p. 335-352.
- (5) VAZQUEZ, Manuela; GALBAN, Carmen. «Perspectivas del empleo de técnicas automatizadas en los estudios bibliométricos». *Actas de II Encuentro hispano-luso de información científica y técnica* (1986), p. 273-280.
- (6) MINISTERIO de Educación y Ciencia; Dirección General de Política Científica: «Recursos humanos en Investigación y Desarrollo». Tomos I y II: «Universidades y CSIC», tomo III: «Centros de Investigación en España» (1986).