

Directorios electrónicos de recursos informativos: aplicación de técnicas bibliométricas para su construcción

JOSÉ NAVARRETE

Universidad de Jaén. Biblioteca

JUAN ANTONIO FERNÁNDEZ LÓPEZ

SAMALY SANTA

Universidad de Granada

El presente estudio intenta aplicar métodos bibliométricos, propios de la gestión y desarrollo de colecciones bibliográficas, en los procesos de recopilación de direcciones de recursos informativos electrónicos distribuidos en la red. El objetivo es dotar al proceso de construcción de los directorios temáticos de metodología que faciliten las tareas de recopilación al mismo tiempo que garanticen calidad en los posibles recursos ofrecidos por un servicio de referencia.

PALABRAS CLAVES: Cibermetría / Bibliometría / Servicio de Referencia / Directorios electrónicos de recursos informativos

ELECTRONIC DIRECTORIES OF INFORMATIVE RESOURCES: APPLICATION OF BIBLIOMETRICS TECHNIQUES FOR ITS CONSTRUCTION.

The present study tries to apply to bibliometrics methods of the management and development of bibliographical collections in the processes of compilation of directions of distributed electronic informative resources in the network. The objective is to equip the process with construction of the thematic directories of methodology that at the same time facilitates the compilation tasks that guarantee quality in the possible resources offered by a service of reference.

KEY WORDS: Cybermetric / Bibliometric / Service of referente / Subjects gateways.

1. INTRODUCCIÓN

La utilización de las nuevas tecnologías de la información en general, e Internet de forma concreta, ha producido un nuevo perfil del servicio de referencia. Prueba de ello, es la actitud de la mayoría de los profesionales a cargo de estos servicios en lo referente a la orientación del usuario en la utilización de estos nuevos recursos. Lo que ha supuesto asumir la responsabilidad de adquirir nuevos métodos de trabajo que han ido permitiendo un mayor conocimiento de todas las herramientas que pueden

utilizar en la recuperación de la información, así como de las técnicas necesarias para la construcción de nuevos instrumentos que posibiliten su difusión.

Entre los nuevos instrumentos de difusión utilizados por los servicios de referencias destacan los directorios electrónicos de recursos informativos externos (DERIE), ubicados, de forma concentrada o dispersa, dentro de los servidores web de las universidades. Por lo general estos directorios se hallan en los espacios web reservados a las bibliotecas universitarias y es responsabilidad de éstas su mantenimiento. De alguna manera, estos instrumentos de difusión suponen la materialización de ideas como la de Fainstein y Mendoza (1), sobre el papel del profesional de la información, dotándolo de un nuevo carácter, al tener que recopilar, organizar y estructurar la información para ponerla a disposición del usuario final. También Magán Wals (2) refuerza esta opinión cuando nos dice que el nuevo profesional estará más volcado en el flujo y transferencia de la información que en su suministro.

En definitiva el bibliotecario de referencia ofrece mejores servicios de información a sus usuarios mediante estas herramientas, su reto actual lo indica Marta Torres (3) será “conocer qué hay en las redes, seleccionar lo pertinente separándolo de lo poco útil, analizar la calidad de lo existente, organizarlo de forma agradable para el usuario y ponerlo a disposición pública”. Esto supone conocer las técnicas necesarias para desarrollar DERIE de calidad.

En los procesos de construcción de los DERIE encontramos dos etapas básicas. La primera relacionada con la localización de recursos en la red potencialmente de interés para una comunidad de usuarios determinada, y la segunda con los procedimientos de catalogación, indización, clasificación y publicación. Sobre esta última etapa, parece haber cada vez más proyectos e investigaciones con resultados más o menos satisfactorios, así lo demuestran proyectos como Dublin Core, CORP, o la utilización de tecnología específica como SFX, así como las nuevas funcionalidades añadidas a los actuales sistemas de gestión bibliotecaria para el tratamiento de recursos electrónicos.

Sin embargo, en lo referente a la identificación y localización de colecciones de enlaces interesantes a recursos informativos en la red, existe la posibilidad de permanecer siempre en la duda sobre la eficiencia y eficacia relacionada con la cobertura y el interés de los recursos que se ofrecen, especialmente si se trata del conjunto de recursos que son gratuitos o de libre acceso. Esta situación es producto en parte a los propios mecanismos que la red nos ofrece para configurar esas colecciones de enlaces de interés, nos referimos a la utilización de buscadores, copiar recursos de DERIE ya existentes o la acumulación de direcciones obtenidas de forma fortuita cuando alguien navega por la red. No obstante, el punto de partida de este estudio se basa en el convencimiento de que estos mecanismos pueden perfeccionarse y enriquecerse.

En este sentido, el presente estudio intenta aplicar métodos bibliométricos, propios de la gestión y desarrollo de colecciones bibliográficas, en los procesos de

recopilación de direcciones de recursos informativos electrónicos distribuidos en la red. El objeto es dotar al proceso de construcción de los DERIE de metodología que facilite las tareas de recopilación al mismo tiempo que garantice calidad en los posibles recursos ofrecidos por un servicio de referencia.

2. MATERIAL Y MÉTODO

Las técnicas metodológicas elegidas para este estudio están basadas en la presunción de un comportamiento, por parte de los contenidos de los *websites* de las bibliotecas, similar al demostrado por Samuel C. Bradford (4) en relación con la productividad de las revistas con respecto a un determinado tema. Creemos que la misma motivación que Bradford tuvo sobre la preocupación acerca de la duplicación de esfuerzos y la insuficiencia de la cobertura de los artículos de revistas científicas por parte de las agencias de indización, es extrapolable a la hora de tomar decisiones en la construcción de un DERIE o evaluar uno ya existente. Sobre esta ley clásica de la Cienciometría incidieron Bar-Ilan (5), quien aplica la Ley de Bradford a los grupos de noticias en un estudio sobre la enfermedad de las vacas locas empleando como método para la recogida de datos el motor de búsqueda AltaVista, llegando a la conclusión de que la distribución de Bradford se puede aplicar a dicho medio e identificar núcleos; o Faba, Guerrero y Moya (6) en un estudio sobre el ajuste de los datos de citas a esta misma distribución, entre otros.

El modelo necesario a considerar para poder hacer la extrapolación del modelo Bradford al conjunto de recursos electrónicos distribuidos en la red sería el siguiente: Imaginemos poder caracterizar cada uno de los espacios de los servidores de las universidades reservados a las bibliotecas con todos los enlaces a recursos externos a su propio dominio a los que hagan referencia. Con ello obtendríamos la idea de que desde los *websites* de las bibliotecas se ofertan recursos, elegidos por bibliotecarios y pensados para un determinado tipo de usuarios, a los que se pueden acceder de forma complementaria a los propios recursos informativos de la biblioteca. En este sentido, podríamos considerar a los servidores referenciados desde los *websites* como las revistas científicas, y al número de referencias (*links*) acumuladas por los mismos servidores, como los artículos especializados en un tema determinado.

En este modelo, también estaríamos asumiendo principios de la teoría de citación, en el entorno web, "*sitations*" (7), ya que el grado de consenso a la hora de referenciar es tomado como indicativo de calidad y pertinencia para incluir a los recursos de dichos servidores dentro de un DERIE. Por otra parte, este indicador de calidad basado en el número de *links* recibidos, es enriquecido por otro indicador que también nos ofrece ponderación relacionada con la calidad de los recursos electrónicos, nos referimos al denominado **factor impacto o de visibilidad** de los servidores Web (**WebIf**) introducido por Ingwersen (8).

$$N^{\circ} Sitations_{Web} = \sum u_{Link}$$

$$WebIf = \frac{\sum x_{Web}}{\sum y_{Web}}$$

Σu_{Link}	Suma de los enlaces a un Servidor desde Universidades distintas
Σx_{Web}	Suma de los enlaces a las páginas de un Servidor desde las Universidades
Σy_{Web}	Suma del número de páginas ofertadas por un Servidor. Estas páginas sólo son contabilizadas si son referenciadas desde los <i>websites</i> de las bibliotecas universitarias. Es necesario especificar, que se ha considerado como recurso electrónico tanto las páginas webs específicas dentro de un servidor, como al servidor propiamente dicho cuando éste es referenciado de forma general.

$$RP_{Web} = N^{\circ} Sitations * WebIf$$

Ambos indicadores caracterizan a los servidores y pueden unirse con el objeto de obtener un único coeficiente que pondere su calidad. El producto de los dos indicadores nos ofrece un tercer indicador que será resumen de los dos anteriores y al que denominaremos **Referenciación Ponderada (RP)**.

De este modo estaremos en condiciones de generar una dispersión por zonas de diferente densidad en cuanto a la concentración del valor de **RP** correspondiente a los servidores y sus recursos informativos ofertados. Es decir, determinaríamos una serie de zonas que recogiendo conjuntos aproximadamente iguales de acumulación de **RP**, necesitarían cantidades crecientes de servidores distintos. Lo relevante es que podemos hacer cálculos aproximativos respecto al número de servidores necesarios para delimitar el conjunto de recursos informativos concretos, que tras un posterior análisis, formarían parte de un DERIE con la garantía de que cubrirían un porcentaje elevado de la demanda.

El banco de prueba para la experimentación del método expuesto, ha sido una base de datos construida con las herramientas y procedimientos de las técnicas de extracción de datos en cibermetría. Con la ayuda del directorio ofrecido por Rebiun (Red de Bibliotecas Universitarias Españolas) se averiguaron las *home pages* de 48 bibliotecas universitarias españolas. A continuación se utilizó la aplicación Xenus para rastrear y capturar todos los enlaces externos a recursos "http" hasta un nivel de profundidad de cinco páginas desde las *home pages*. Posteriormente se diseñó una base de datos relacional con Microsoft Acces donde se cargaron todas las urls capturadas, las cuales fueron sometidas a un proceso de depuración, mediante la creación de nuevas tablas y consultas, para descartar duplicados y falsos recursos externos y así obtener las relaciones de frecuencias correctas (Tabla I).

TABLA I
DATOS DE VOLCADO DE INFORMACIÓN

• Home Pages de bibliotecas analizadas	48
• N° de Recursos externos capturados	22.082
• N° de Servidores referenciados	10.650

Para obtener el número necesario de servidores que ha proporcionado el conjunto de recursos que conformaría el DERIE, se ha recurrido a la acumulación de los servidores necesarios para cubrir no menos del 33 % de las referencias ponderadas (**RP**) totales, es decir se han seleccionado los servidores integrantes de la zona nuclear de Bradford[i].

La cuestión de la determinación del núcleo y las zonas sucesivas mediante un ajuste fraccionado de modelos matemáticos (potencial para el núcleo, parte curva de la gráfica, y logarítmico para las zonas sucesivas, parte recta de la gráfica (9)) nos parece de una complejidad innecesaria para los objetivos que se persiguen en este estudio. Por ello consideramos como válido el procedimiento de dividir el valor total del **RP** en tres conjuntos aproximadamente iguales y a partir de aquí identificar un núcleo y varias zonas de densidad decreciente. Este procedimiento es impugnado por autores como Ferreiro (10) por no ser demasiado estricto matemáticamente desde el punto de vista del postulado verbal Bradfordiano. Sin embargo, creemos que para lo pretendido en este análisis, se puede sacrificar la exactitud en la metodología de los cálculos.

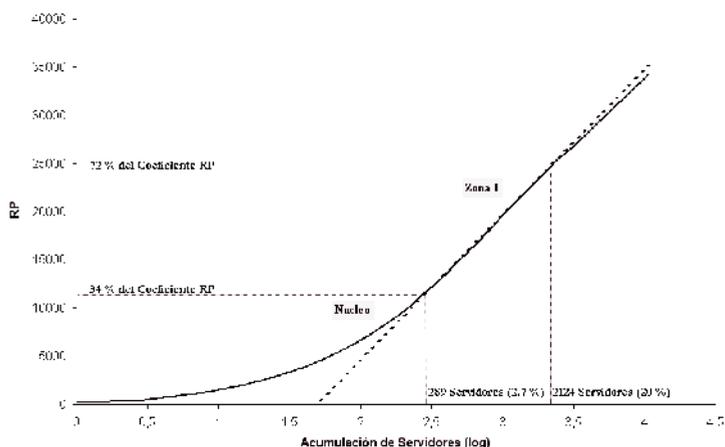
3. RESULTADOS

Tras asumir los principios metodológicos anteriormente descritos obtenemos un ranking de servidores donde se han valorado los dos aspectos cualitativos a tener en cuenta a la hora de seleccionar servidores de interés, nos referimos a la visibilidad y al consenso de la mayoría de las bibliotecas universitarias españolas en referenciarlos. La siguiente etapa pasa por ver el comportamiento de estos servidores en una distribución Bradford con el fin de observar el grado de concentración del indicador elegido como ponderador de los servidores. Esto queda explicado en las frecuencias en la distribución de forma acumulada (Tabla II) y su correspondiente representación gráfica en una Curva Bradford (Figura 1).

i. Pueden consultarse los servidores en:

http://sicareport.ugr.es:8080/sica2003/menu_presentacion/anexo_II.htm

FIGURA 1
DISTRIBUCIÓN BRADFORD
SERVIDORES / REFERENCIACIÓN PONDERADA (RP)



En la Tabla II, la primera columna tiene el número de rango r en la ordenación secuencial descendente de los servidores por coeficiente de RP . La segunda (variable x) indica la cantidad de servidores con igual coeficiente RP ; la tercera columna (variable y) el propio valor del coeficiente RP . Seguidamente las columnas correspondientes a Σx y Σy , son las acumulaciones de las dos variable anteriores, y por último, la sexta columna indica el logaritmo de los valores acumulados por Σx , necesario para la representación de la curva de Bradford. Los resultados obtenidos de la distribución son los siguientes:

Zona nuclear:	constituida por 289 servidores (2,7%) que concentran 11659 de RP (34%)
Zona 1º:	constituida por 2124 servidores (20%) que concentran 24529 de RP (71,5%)
Nº de servidores seleccionados:	289 servidores (2,7%) que concentran 11659 de RP (34%)

A continuación se exponen las características de forma comparativa entre el grupo de servidores pertenecientes al núcleo y el resto de la distribución.

	Zona Nuclear	Resto de la Distribución
Rango de RP	20 - 173	1 - 19
Rango de Situations	5 - 41	1 - 18
Rango de Weblf	1,06 - 13	1 - 8

Tabla II
DISTRIBUCIÓN BRADFORD
SERVIDORES / REFERENCIACIÓN PONDERADA

r	x	y	Σx	Σy	Log Σx	r	x	y	Σx	Σy	Log Σx
1	1	173	1	173	0.000	43	4	42	81	5906	1.908
2	2	171	3	515	0.477	44	3	41	84	6029	1.924
3	2	169	5	853	0.699	45	7	40	91	6309	1.959
4	1	144	6	997	0.778	46	5	39	96	6504	1.982
5	1	130	7	1127	0.845	47	2	38	98	6580	1.991
6	1	127	8	1254	0.903	48	2	37	100	6654	2.000
7	1	121	9	1375	0.954	49	15	36	115	7194	2.061
8	1	116	10	1491	1.000	50	4	35	119	7334	2.076
9	1	114	11	1605	1.041	51	2	34	121	7402	2.083
10	1	110	12	1715	1.079	52	7	33	128	7633	2.107
11	1	109	13	1824	1.114	53	7	32	135	7857	2.130
12	1	100	14	1924	1.146	54	5	31	140	8012	2.146
13	1	96	15	2020	1.176	55	6	30	146	8192	2.164
14	2	92	17	2204	1.230	56	3	29	149	8279	2.173
15	2	91	19	2386	1.279	57	11	28	160	8587	2.204
16	2	84	21	2554	1.322	58	14	27	174	8965	2.241
17	4	81	25	2878	1.398	59	7	26	181	9147	2.258
18	1	80	26	2958	1.415	60	44	25	225	10247	2.352
19	1	76	27	3034	1.431	61	20	24	245	10727	2.389
20	2	72	29	3178	1.462	62	6	23	251	10865	2.400
21	1	68	30	3246	1.477	63	11	22	262	11107	2.418
22	1	67	31	3313	1.491	64	12	21	274	11359	2.438
23	1	66	32	3379	1.505	65	15	20	289	11659	2.461
24	2	64	34	3507	1.531	66	17	19	306	11982	2.486
25	1	63	35	3570	1.544	67	30	18	336	12522	2.526
26	1	62	36	3632	1.556	68	7	17	343	12641	2.535
27	2	61	38	3754	1.580	69	91	16	434	14097	2.637
28	3	60	41	3934	1.613	70	14	15	448	14307	2.651
29	1	59	42	3993	1.623	71	29	14	477	14713	2.679
30	2	58	44	4109	1.643	72	26	13	503	15051	2.702
31	1	57	45	4166	1.653	73	61	12	564	15783	2.751
32	2	56	47	4278	1.672	74	19	11	583	15992	2.766
33	3	55	50	4443	1.699	75	54	10	637	16532	2.804
34	2	53	52	4549	1.716	76	216	9	853	18476	2.931
35	2	51	54	4651	1.732	77	128	8	981	19500	2.992
36	1	50	55	4701	1.740	78	56	7	1037	19892	3.016
37	11	49	66	5240	1.820	79	95	6	1132	20462	3.054
38	2	48	68	5336	1.833	80	99	5	1231	20957	3.090
39	1	47	69	5383	1.839	81	893	4	2124	24529	3.327
40	1	46	70	5429	1.845	82	252	3	2376	25285	3.376
41	4	45	74	5609	1.869	83	737	2	3113	26759	3.493
42	3	43	77	5738	1.886	84	7537	1	10650	34296	4.027

Filtrados los servidores de la zona nuclear se ha pasado al análisis de los recursos concretos ofertados. Para ello se ha recurrido al establecimiento de un ranking por *nº de Sitation* y siguiendo esta frecuencia se han establecido dos grupos de recursos divididos por el valor de corte de *Sitation* de la zona nuclear de servidores (primer grupo: ≥ 5 *Sitation*; segundo grupo: ≤ 4 *Sitation*). Seguidamente se ha procedido al análisis individual de cada uno de los recursos del primer grupo, considerando a éste como el principal conjunto seleccionado. Aunque finalmente se decide ampliar el margen establecido por el umbral ≥ 5 con el conjunto adicional de recursos de una frecuencia igual a 4, es decir un punto por debajo de lo establecido en primer momento. Este hecho podría servir de garantía para que el margen de error cometido fuese por exceso. Durante este análisis se ha añadido a cada recurso una denominación identificativa de su contenido y una clasificación basada en tipologías de fuentes de información. Finalmente se han obtenido un total de **442 recursos** [ii] diferentes.

Total de recursos (páginas) acumulados por los 289 servidores nucleares:	3251
Grupo primero de recursos ≥ 5 :	353
Grupo segundo de recursos ≤ 4 :	2898
Conjunto adicional de recursos =4:	89

El producto final queda configurado por un conjunto de páginas html generadas de forma automática a partir de nuestra base de datos, dando como resultado una interfaz de navegación que atiende a la siguiente estructura: directorios; información de carácter oficial; información bibliográfica; y referencia general. A su vez estas categorías son subdivididas en otras más específicas y al mismo tiempo más dinámicas, ya que estarán sometidas a continuos cambios durante el mantenimiento del DERIE (Figura 2). De forma más detallada en las tablas III y IV se puede observar la distribución por tipos de fuentes de los recursos.

Tabla III
DISTRIBUCIÓN DE LOS RECURSOS EN CATEGORÍAS GENERALES

Clasificación	Nº de recursos
Información bibliográfica	171
Directorios	93
Información de carácter oficial	79
Referencia general	66

ii. Pueden consultarse los recursos en:

http://sicareport.ugr.es:8080/sica2003/menu_presentacion/anexo_III.htm

Tabla IV
DISTRIBUCIÓN DE LOS RECURSOS EN CATEGORÍAS ESPECÍFICAS

Subclasificación	Nº de recursos
Bibliotecas más visitadas	49
Directorios institucionales	48
Enlaces a recursos de biblioteconomía	41
Buscadores	35
Organismos oficiales más visitados	31
Editoriales y librerías más visitadas	30
Organismos e instituciones relacionadas con la ciencia y tecnología	28
Bases de datos más visitadas	24
Prensa	24
Directorios de bibliotecas y archivos	21
Catálogos colectivos	18
Revistas electrónicas	16
Publicaciones oficiales	13
Directorios temáticos de recursos electrónicos	12
Diccionarios, enciclopedias, etc...	9
Fuentes estadísticas	4
Medios de comunicación	2
Becas y convocatorias	2
Listas de distribución	1

4. CONCLUSIONES

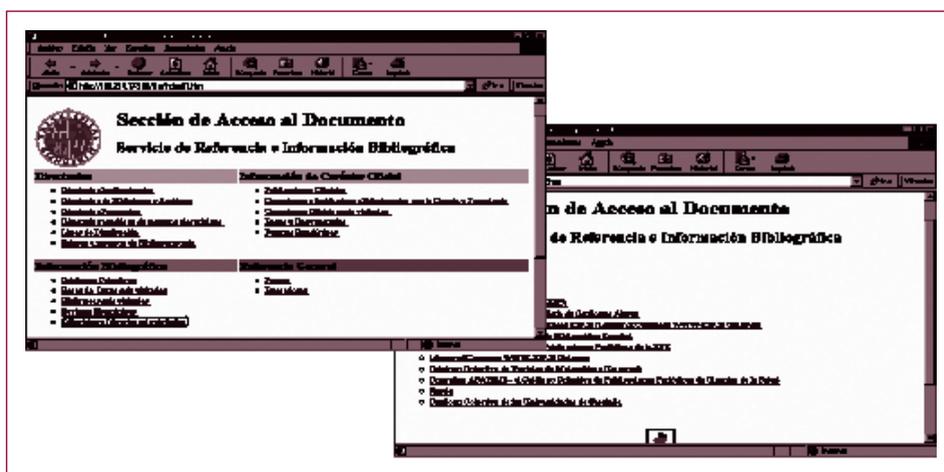
Tras el análisis de los resultados obtenidos, llegamos a las siguientes conclusiones:

1. Observado el grado de aplicabilidad de la ley de Bradford en la distribución de los recursos electrónicos en Internet, podemos considerar la aplicación del modelo Bradford válido como método de ponderación a la hora de seleccionar recursos electrónicos.
2. El resultado final obtenido en este estudio concreto también refleja una práctica del método para el análisis de conductas o patrones a la hora de estudiar los contenidos de los *webites* de las bibliotecas.
3. Creemos que la muestra sobre la que se aplica nuestro método, debería ser complementada. Sugerimos realizar estudios con la misma metodología pero

con otra selección de recursos. Ésta podría ser una selección de servidores especializados por temas, ya que en las bibliotecas universitarias abundan los recursos de carácter general, al menos entre el conjunto de enlaces gratuitos.

4. El método presentado para la selección de recursos en la construcción de DERIE debe ser utilizado de forma conjunta con los otros métodos tradicionales a la hora de la elección de recursos (buscadores y consulta de *Subjects gateways* especializados), ya que nuestro método refleja la calidad de los recursos de forma indirecta; es el consenso lo que representa la calidad del recurso. Nuestra selección de recursos, puede que no esté integrado por las mejores páginas, desde un punto de vista objetivo de calidad, pero sí son las más usadas.

Figura 2
ASPECTO ESTÉTICO DEL DERIE



REFERENCIAS

- (1) FAINSTEIN LAMUERDA, G.; MENDOZA GARCÍA, I. “El impacto de Internet en el trabajo profesional del documentalista : contrastes y perspectivas : referencias a casos concretos de centros de documentación y bibliotecas”. En *Actas de las V Jornadas Españolas de Documentación Automatizada (Cáceres, 1996)*, pp. 269-286.
- (2) MAGAN WALS, J. A. “Los servicios de información y referencia bibliográfica. Situación actual y aprovechamiento de los recursos”. En: Magan Wals (coord.). *Tratado básico de biblioteconomía*. Madrid: Editorial Complutense, 1996, pp. 341-365.
- (3) TORRES, M. El Impacto de las Autopistas de la Información sobre la Comunidad Académica y los Bibliotecarios. Cursos de Verano de la Universidad Complutense de Madrid sobre Las Autopistas de la Información: el Reto del Siglo XXI. 1995, julio, p.75.
- (4) BRADFORD, S. C. Sources of information on specific subjects. *Engineering*, 1934, vol. 137 (8), 5-6.
- (5) BAR-ILAN, J. The “mad cow disease”, usenet groups and bibliometric laws. *Scientometrics*, 1997, vol. 39 (1), 29-55.
- (6) FABA, C.; GUERRERO, V. P.; MOYA, F. (2003) “Sitation” distributions and Bradford’s law in a closed web space. *Journal of Documentation*, 2003, vol. 59 (5), 558-580.
- (7) MCKIERNAN, G. CitedSites(sm): *Citation Indexing of Web resources* [en línea]. [Consulta: 17 de agosto de 2006]. Disponible en: <http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm>
- (8) INGWERSEN, P. The calculation of Web impact factors. *Journal of Documentation*, 1998, vol. 54 (2), 236-243.
- (9) BROOKES, B. C. Bradford’s law and the bibliography of science. *Nature*, 1969, vol. 224, 953-956.
- (10) FERREIRO ALAEZ, L. *Bibliometría (análisis bivariante)*. Madrid : EYPASA, 1993