



Facultad de  
**Comunicación y Documentación**

UNIVERSIDAD DE GRANADA

GRADO EN INFORMACIÓN Y DOCUMENTACIÓN

TRABAJO FIN DE GRADO

**Análisis de la Producción Científica en Recuperación de  
Información en Library & Information Science : Web of  
Science 1990-2020**

Presentado por:

**D<sup>a</sup>. Apolonia Millán Castro**

Tutor: D. Benjamín Vargas Quesada

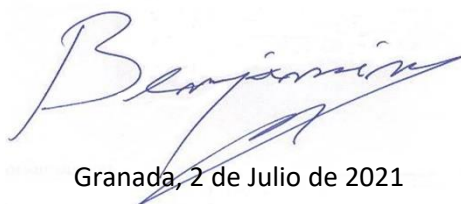
Curso académico 2020/2021



D. Benjamín Vargas Quesada tutor/a del trabajo titulado **Análisis de la Producción Científica en Recuperación de Información en Library & Information Science: Web of Science 1990-2020**

realizado por la alumna **Apolonia Millán Castro**, INFORMA que dicho trabajo cumple con los requisitos exigidos por el Reglamento sobre Trabajos Fin del Grado en Información y Documentación para su defensa.

3



Granada, 2 de Julio de 2021

Fdo.: Benjamín Vargas Qusada

Por la presente dejo constancia de ser el/la autor/a del trabajo titulado **Análisis de la Producción Científica en Recuperación de Información en Library & Information Science : Web of Science 1990-2020**

que presento para la materia Trabajo Fin de Grado del Grado de Información y Documentación tutorizado por el/la profesor/a **Vargas Quesada, Benjamín** durante el curso académico 20- 21.

4

Asumo la originalidad del trabajo y declaro que no he utilizado fuentes (tablas, textos, imágenes, medios audiovisuales, datos y software) sin citar debidamente, quedando la Facultad de Comunicación y Documentación de la Universidad de Granada exenta de toda obligación al respecto.

Autorizo a la Facultad de Comunicación y Documentación a utilizar este material para ser consultado con fines docentes dado que constituyen ejercicios académicos de uso interno.

  02   /   07   /  2021 

Fecha Firma Apolonia Millán



**Análisis de la Producción Científica en Recuperación de Información en Library & Information Science : Web of Science 1990-2020**

**INDICE**

**1. INTRODUCCIÓN**

**6**

**2. OBJETIVOS**

**8**

**3. MATERIAL Y METODOLOGÍA**

**9**

**4. CONCLUSIONES**

**55**

**BIBLIOGRAFIA**

**56**

# **Análisis de la Producción Científica en Recuperación de Información en Library & Information Science : Web of Science 1990-2020**

## **RESUMEN**

Cada vez, son más los usuarios que se acercan a los centros de documentación con la necesidad de cubrir una necesidad concreta de información. Surge la duda de cómo hacerles llegar un conjunto de documentos relevantes que satisfaga esa necesidad concreta de información cuando lo solicitan con la seguridad de que se le está dando información pertinente y relevante.

Con el presente trabajo se pretende analizar cómo ha evolucionado el área de la Recuperación de la Información a nivel mundial desde la década de los 90 hasta el año 2020. Y como con un software denominado Biblioshiny ofrecer esa información al usuario es garantía de ofrecerle una respuesta adecuada a su consulta.

**Palabras Clave:** Recuperación de Información, documentos, usuarios, información consulta, Biblioshiny

## **ABSTRACT**

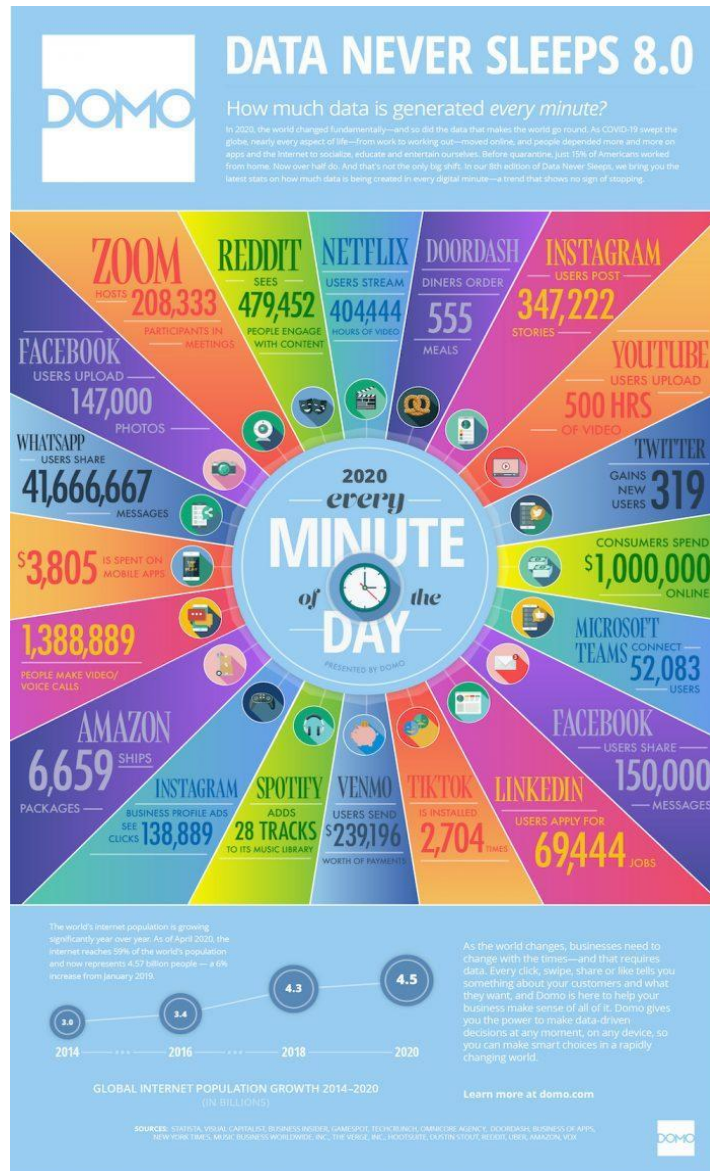
More and more users are approaching documentation centers with the need to meet a specific need for information. The question arises as to how to send them a set of relevant documents that satisfies that specific need for information when they request it with the assurance that pertinent and relevant information is being given.

The present work aims to analyze how the area of Information Retrieval has evolved worldwide from the 90s to 2020. And as with a software called Biblioshiny, offering that information to the user is a guarantee of offering an answer appropriate to your query.

**Keywords:** Information Retrieval, documents, users, query information, Biblioshiny

# 1. INTRODUCCIÓN

Según el informe Data Never Sleeps, elaborado por el sistema operativo basado en la nube Domo, cada día se crean en Internet más de 205 billones de bytes de datos y esta cantidad sigue aumentando. (Domo org., 2020)



## Informe Data Never Sleep

De hecho, cualquier niño pequeño con un móvil en la mano tiene acceso inmediato a más información que la que tenía el presidente de los Estados Unidos Ronald Reagan en los años 80 durante su mandato.

Esto hace que nos preguntemos si toda esta información es útil. Y más allá de esto, cuando surge una necesidad de información concreta de un usuario cómo hacer para que la salida de un conjunto de documentos satisfaga potencialmente esta necesidad.

Esta es la función de un sistema de Recuperación de Información (IR) o Information Retrieval, indicar al usuario qué documentos son potencialmente relevantes para la necesidad de información que le surja.(Baeza-Yates & Ribeiro-Neto, 1999)

La recuperación de la Información o IR, es el área de la ciencia y la tecnología que trata de la adquisición, representación, almacenamiento, organización y acceso a elementos de información.(Vilares, 2019 )

Desde el punto de vista de un servicio de Información, ya sea biblioteca, archivo..., dada una necesidad de información, un proceso de IR ofrece un conjunto de documentos que satisfacen la necesidad de información que ha surgido. Esto es importante, ya que un IR no tiene la función de dar al usuario la información deseada, sino la de indicarle qué documentos pueden servirle para cubrir su necesidad de información.

## **1.1 TERMINOLOGÍA BÁSICA**

Es interesante introducir algunas definiciones de ciertos términos de los que se esperan distintas cosas si hablamos de recuperación de la Información.

Cuando hablamos de documento en RI el término se refiere, de forma genérica, a la unidad almacenada por el sistema y disponible para su consulta y recuperación. Por ejemplo, un documento sería un capítulo de un libro.

La necesidad de información del usuario, expresada en términos que el sistema pueda comprender se denomina query o consulta. Los resultados que esta ofrece, se ofrecen ordenados por grado de similaridad o relevancia.(Rijsbergen, 1999)

Relevancia es un término importante y que puede confundir, ya que no se refiere a la relevancia en sentido estricto, sino más bien a la impresión que tendrá el usuario cuando vea el documento devuelto y la necesidad de información expresada. Es un concepto muy subjetivo.



## 2. OBJETIVOS

### 2.1.1 OBJETIVO GENERAL

El objetivo principal de este trabajo es hacer un análisis bibliométrico básico de la materia de recuperación de la información a nivel mundial y en los últimos años, reflejando la evolución de la investigación y lo que se ha publicado en este campo. Para ello se ha hecho una búsqueda en Web of Science abarcando los años 1990-2020.

### 2.1.2. OBJETIVOS ESPECÍFICOS

Además de esta idea general se definen una serie de objetivos específicos descritos en el programa Biblioshiny que se enumeran a continuación como perspectivas a tener en cuenta para conocer qué se ha publicado acerca de la Recuperación de la información en los últimos años.

Estos objetivos son:

Cotejar la base de datos de la que extraemos la información. En este apartado se consideran distintos subaspectos como la Información principal o Main Information, la producción científica anual, o el promedio de citas por año.

Analizar las fuentes de las que extraemos la información. Se verán los recursos más relevantes, los recursos más citados localmente, y una referencia a la ley bibliométrica de Bradford.

Visibilizar a los autores más relevantes, más citados y con más producción en esta materia.

Conocer los documentos: en este apartado se enumeran los documentos más citados a nivel local y global así como las palabras más utilizadas en lo referente a nuestra búsqueda.

Comprender la estructura conceptual, refiriéndose a la co-occurrence en la red, los mapas temáticos que surgen y la evolución de la materia objeto de investigación.

Describir la concitación en la red y la historiografía del tema objeto de estudio.

Mostrar la estructura social que permite conocer la colaboración entre autores en lo que tiene que ver con la materia objeto de investigación.

## 3. MATERIAL Y METODOLOGÍA

### 3.1.1. MATERIAL

Para llevar a cabo el siguiente análisis bibliométrico se ha utilizado el software gratuito Biblioshiny.

Este software es una aplicación para bibliometría que corre sobre R y sirve para hacer análisis e informes bibliométricos sobre información en distintas bases de datos.

Es una herramienta flexible que facilita la integración con otros paquetes estadísticos y gráficos. Combina la funcionalidad de paquetes bibliométricos con la facilidad de uso de aplicaciones web usando el entorno de Shiny.

Además, en este trabajo se ha utilizado como fuente de información la base de datos internacional Web of Science (WOS) de Clarivate Analytics. Se ha escogido esta base de datos porque permite su procesamiento en el programa RStudio y esto facilita la labor de preparar los datos para analizarlos.

De ella se han recogido los datos empleados en el estudio de la materia a tratar.

### 3.1.2. METODOLOGÍA

Dentro de la base de datos WOS se ha hecho la búsqueda la base de datos conocida como Colección principal de Web of Science.

La ecuación de búsqueda utilizada ha sido la siguiente:

Etiqueta de campo AK= (palabras clave del autor). Se escoge esta etiqueta de campo porque nadie mejor que el autor para describir de qué trata su trabajo.

Etiqueta de campo KP= (keyword plus). Esta etiqueta se elige porque con técnicas de procesamiento de lenguaje natural analiza las palabras del título y del resumen y extrae aquellas que pueda considerar que son más representativas. Complementa a los author keywords.

Por lo tanto, la ecuación de búsqueda queda de la siguiente manera:

AK= (“Information retrieval”) OR KP=(“Information retrieval”)

Además se limita la búsqueda a los años 1990 hasta 2020. Y se limita al campo de la Información y Documentación.

En lo relacionado a los tipos de documentos en los que se publica, se decide que es interesante saber que se ha publicado en todos los tipos de documentos, y no limitar esta búsqueda a artículos científicos solamente.

Se descargan los registros seleccionados como registros completos y Referencias citadas y en cuanto al formato de archivo se elige texto sin formato. Esto se hace así porque este tipo de formato nos permite ver todo el texto descargado con etiquetas y en el bloc de Notas mediante líneas. Así es más fácil averiguar si hay algún error, dónde se encuentra.

A continuación, de acuerdo con los objetivos del análisis que se pretende realizar, se descargaron los datos de la base de datos Web of Science (WOS), para poder trabajar con ellos y procesarlos en el software Biblioshiny.

Una vez que tenemos los datos, comienza el trabajo en Biblioshiny.(Aria, M., & Cuccurullo, C., 2017)

## 3.2. Información sobre el Conjunto de Datos

Todos los datos, gráficos y tablas que se citan a continuación han sido extraídos directamente del software Biblioshiny. Esta herramienta ha sido utilizada para hacer el informe bibliométrico. (Aria, M., & Cuccurullo, C., 2017)

### 3.2.1 Información Principal

DESCRIPCIÓN	RESULTADOS
Timespan	1990:2020
Sources (Journals, Books, etc)	282
Documents	2554
References	57721
DOCUMENT TYPES	
article	1839
article; proceedings paper	64
book review	194
editorial material	26
proceedings paper	301
review	79
DOCUMENT CONTENTS	
Keywords Plus (ID)	1511
Author's Keywords (DE)	4395
AUTHORS	
Authors	4187
Author Appearances	5881
Authors of single-authored documents	648

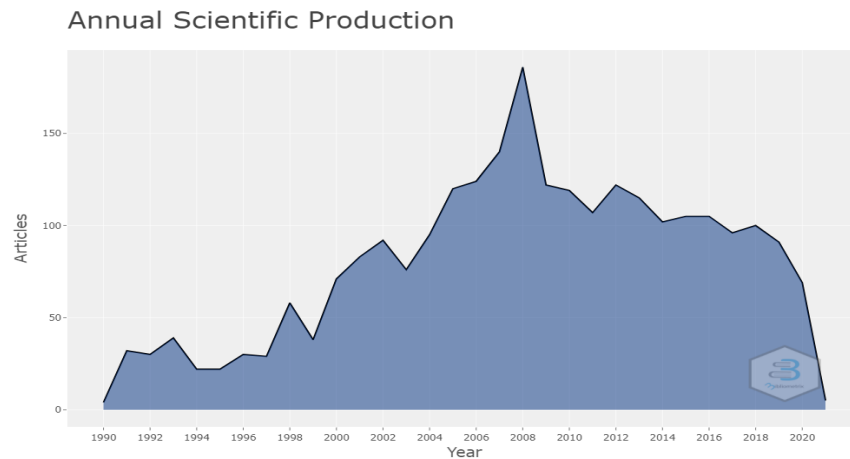
Tabla 1. Información bibliométrica

Esta tabla muestra cuáles son los datos que vamos a analizar. El periodo de años que hemos seleccionado en WOS ha sido desde el año 1990 hasta el año 2020. El total de documentos extraído para su análisis ha sido de 2554 documentos descargados.

En cuanto a la tipología de documentos a analizar, decir que es diversa. Se analizarán artículos, revistas, participaciones en congresos..

### 3.2.2 Producción Científica por Años

Con la ayuda del software Biblioshiny se analiza cómo ha sido la producción científica por años.



Año	Artículos
1990	4
1991	32
1992	30
1993	39
1994	22
1995	22
1996	30
1997	29
1998	58
1999	38
2000	71
2001	83
2002	92
2003	76
2004	95
2005	120
2006	124
2007	140
2008	186
2009	122

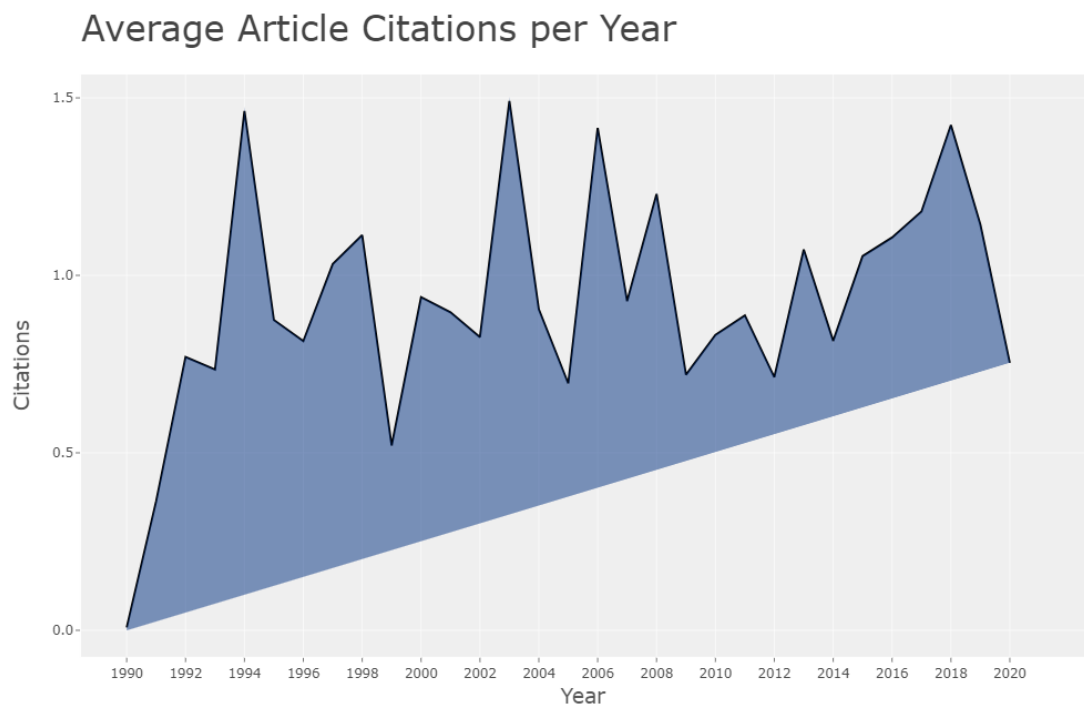
2010	119
2011	107
2012	122
2013	115
2014	102
2015	105
2016	105
2017	96
2018	100
2019	91
2020	69

Figura 1. Evolución de la Producción Científica por años

En lo relacionado a la evolución de la producción científica por años se observa que en el primer año analizado, el año 1990, la producción es de 4 documentos. Este número se eleva a 32 documentos en el año 1991. Pero el pico de producción se observa en el año 2008. Estos picos implican que se ha descubierto algo y se empieza a publicar mucho acerca de ello. Por lo tanto se puede afirmar que en el año 1998 se tiende a publicar acerca de la recuperación de la información, es decir, hay una nueva tendencia o surge algo que hace que se investigue mucho y se publique mucho sobre recuperación de información.

Si observamos la información que se publica ese año descubriremos que la mayoría tiene que ver con la visualización de la información. Efectivamente, a partir del año 2008 se empieza a dar importancia a este tema y como bien nos indica el software utilizado hay un repunte importante en lo que tiene que ver con la importancia de la recuperación de la información para tener visualización como investigadores.

### 3.2.3 Proporción de citas por año



Año	Docs	Media de Citas por Art.	Media de Citas por Año	Años Citables
1990	4	0,25	0,01	31
1991	32	10,91	0,36	30
1992	30	22,33	0,77	29
1993	39	20,56	0,73	28
1994	22	39,50	1,46	27
1995	22	22,73	0,87	26
1996	30	20,37	0,81	25
1997	29	24,76	1,03	24
1998	58	25,60	1,11	23
1999	38	11,45	0,52	22
2000	71	19,70	0,94	21
2001	83	17,92	0,90	20
2002	92	15,70	0,83	19
2003	76	26,84	1,49	18

2004	95	15,38	0,90	17
2005	120	11,13	0,70	16
2006	124	21,23	1,42	15
2007	140	12,98	0,93	14
2008	186	15,98	1,23	13
2009	122	8,64	0,72	12
2010	119	9,15	0,83	11
2011	107	8,87	0,89	10
2012	122	6,42	0,71	9
2013	115	8,58	1,07	8
2014	102	5,71	0,82	7
2015	105	6,32	1,05	6
2016	105	5,53	1,11	5
2017	96	4,72	1,18	4
2018	100	4,27	1,42	3
2019	91	2,29	1,14	2
2020	69	0,75	0,75	1

Figura 2. Promedio de Citas por Año

En cuanto al promedio de citas por año se observa que hay años en los que lo que se produce es muy utilizado para investigar. Esto se puede observar en los picos de la figura 2. En los años 1994 la media de citas por año es 1.5 y aunque parece baja es una subida respecto a lo anterior. Igual sucede con los años 2004 y 2008 donde la media de citas aumenta. Esto es importante ya que conociendo los artículos más citados en esos años, sabremos las líneas de investigación que se están en auge en ese momento.

Se observa que la revista con más citas es Journal of America Source Information Science, le sigue Information process and management.

Es un dato a tener en cuenta que Journal of America Source Information Science es la misma revista que Journal of american society for Information Science and Technology.

Por lo tanto, la suma de los dos resultados del número de citas la hacen aún más citada.



### 3.2.4 Gráfico de tres dimensiones: keywords, autores, documentos fuente.

Esta perspectiva nos muestra la representación por los campos keywords, autores y revistas.

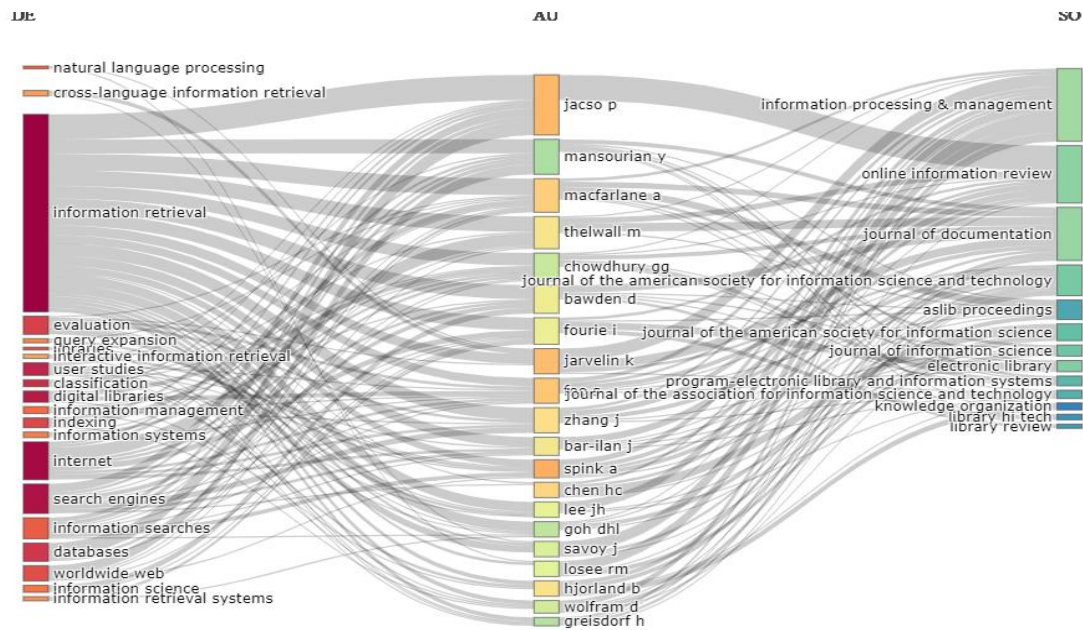


Figura 3. Gráfico de 3 dimensiones

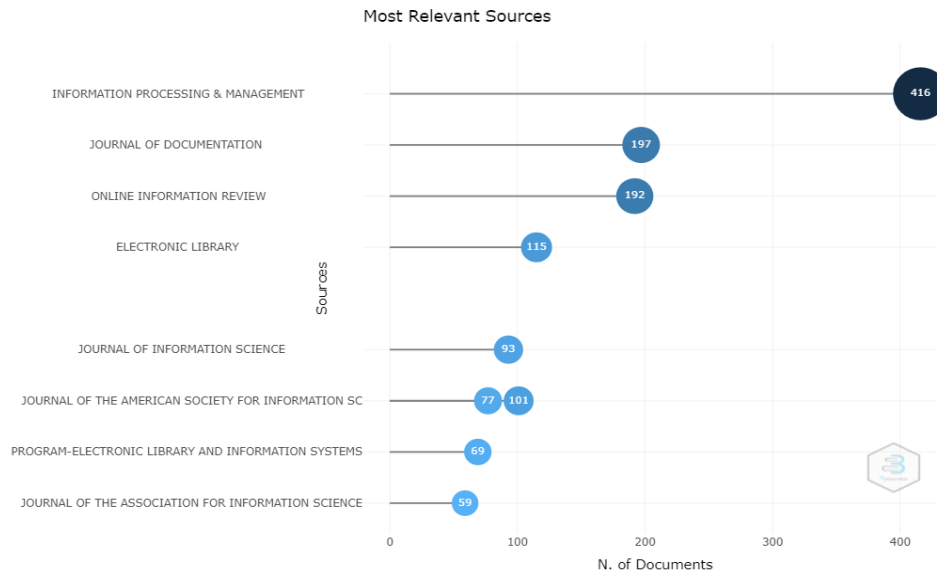
Biblioshiny, representa un gráfico en tres dimensiones con la información que necesitamos. Esta perspectiva nos muestra la representación por los campos keywords, autores y revistas. Nos enseña la relación entre las principales keywords que se están utilizando en el campo de RI, quiénes son los autores que están utilizando esas keywords y en qué revistas publican.

Podemos observar que el país que más publica acerca de Information Retrieval es Estados Unidos. Si se analiza España se observa que también se publica bastante acerca de Recuperación de la Información (IR), pero Estados Unidos le gana.

La revista dónde más se publica acerca de este tema según nos muestra el gráfico es Information processing & management. Esto resulta paradójico ya que esta revista publica investigaciones vanguardistas que se llevan a cabo en el campo de la Informática junto con las Ciencias de la Información relacionadas con la publicidad, el marketing tecnológico y los negocios. Sin embargo, son los que más publican sobre IR.

### 3.3. Documentos Fuente

#### 3.3.1 Revistas con Mayor Producción Científica



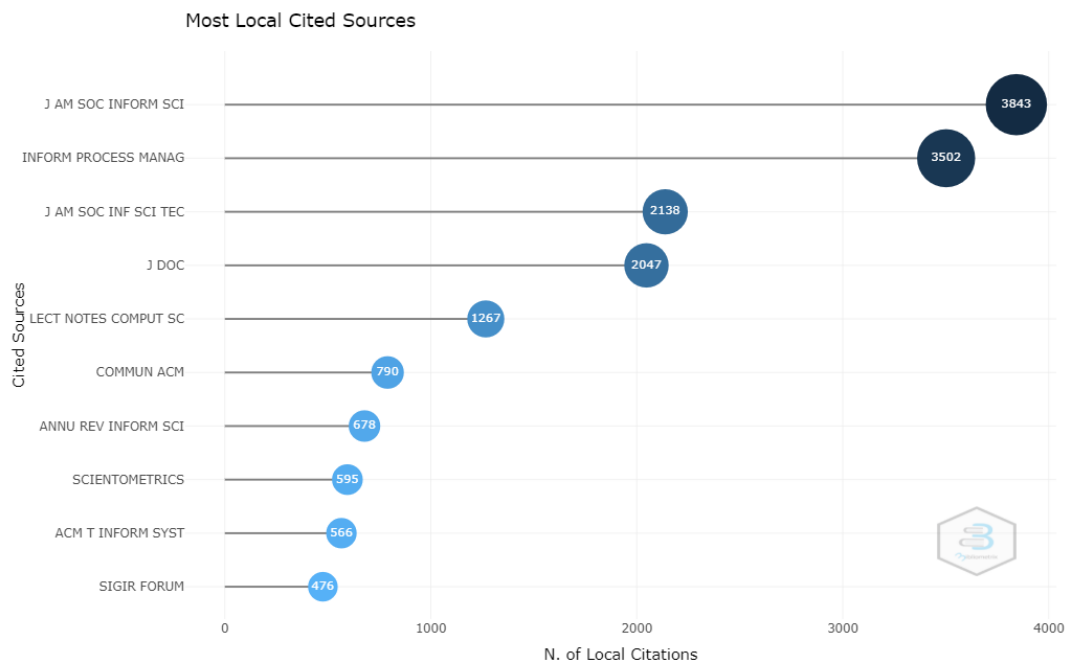
Sources	Articles
INFORMATION PROCESSING & MANAGEMENT	416
JOURNAL OF DOCUMENTATION	197
ONLINE INFORMATION REVIEW	192
ELECTRONIC LIBRARY	115
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIE	101
JOURNAL OF INFORMATION SCIENCE	93
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIE	77
PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS	69

Figura 3. Relación revistas con mayor producción y artículos publicados

En lo relacionado con las fuentes de las que se toma la información, la figura 3 muestra cuales son las revistas con mayor producción científica de entre las extraídas de WOS. Observamos de nuevo que Information Processing & Management es la revista que más ha publicado acerca de RI, con un total de 416 artículos. Ya se comentó que es paradójico que una revista que tiene que ver con el mundo de las finanzas, el marketing y la publicidad publique tal cantidad de artículos relacionados con RI, pero no olvidemos que sus investigaciones están relacionadas con el área de la Informática junto con las Ciencias de la Información.

A esta le sigue con 197 artículos Journal of Documentation, Online Information Review con 192 artículos y en las siguientes se observa un descenso de artículos escritos acerca de IR hasta los 69 extraídos de la revista Program-Electronic Library and Information Systems.

### 3.3.2 Revistas más Citadas



FUENTES	ARTICULOS
J AM SOC INFORM SCI	3843
INFORM PROCESS MANAG	3502
J AM SOC INF SCI TEC	2138
J DOC	2047
LECT NOTES COMPUT SC	1267
COMMUN ACM	790
ANNU REV INFORM SCI	678
SCIENTOMETRICS	595
ACM T INFORM SYST	566
SIGIR FORUM	476

Figura 4. Revistas más citadas

En cuanto a la perspectiva de Revistas más Citadas, desde el punto de vista del análisis de la recuperación de la información, las revistas que tienen los documentos más citados son las revistas más importantes.

El objetivo de esta perspectiva es saber cuáles son las revistas que más se utilizan en un centro de información, que serán a su vez las más utilizadas por los usuarios.

En caso de querer hacer un análisis del desarrollo de la colección, esta perspectiva sirve para saber cuáles son las más utilizadas.

Desde el punto de vista del análisis de la recuperación de la información, las revistas que tienen los documentos más citados son las revistas más importantes.

El objetivo de esta perspectiva es saber cuáles son las revistas que más se utilizan en un centro de información, cuyo conocimiento es el que más se usa por parte de los usuarios.

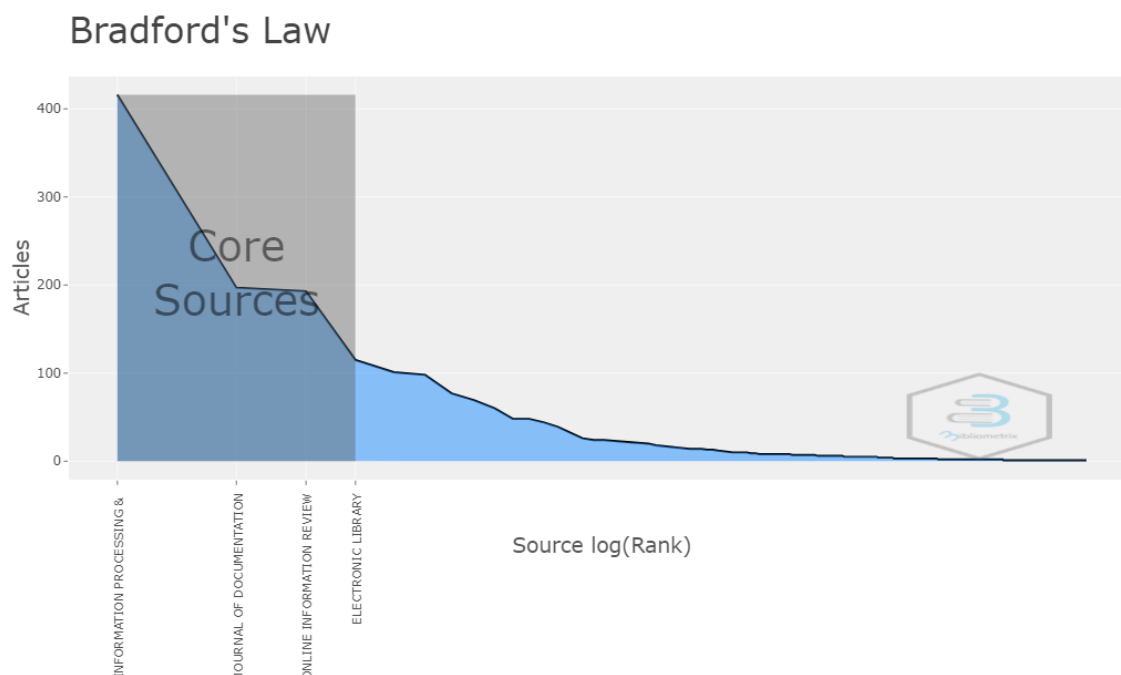
En caso de querer hacer un análisis del desarrollo de la colección, esta perspectiva sirve para saber cuáles son las más utilizadas.

Se observa que la revista con más citas es Journal of America Source Information Science, le sigue Information process and management.

Un dato a tener en cuenta que Journal of America Source Information Science es la misma revista que Journal of american society for Information Science and Technology, pero con el título distinto porque en su momento lo decidieron así las editoriales.

Por lo tanto, la suma de los dos resultados del número de citas la hacen aún más citada.

### 3.3.3. Revistas Nucleares (Ley de Bradford)



REVISTAS	Orden	Frecuencia	Frecuencia	Zona
INFORMATION PROCESSING & MANAGEMENT	1	416	416	Zone 1
JOURNAL OF DOCUMENTATION	2	197	613	Zone 1
ONLINE INFORMATION REVIEW	3	193	806	Zone 1
ELECTRONIC LIBRARY	4	115	921	Zone 1

Figura 5. Revistas Nucleicas

No podemos olvidar citar a Bradford y su ley bibliométrica. La ley de Bradford dice que si las revistas se organizan en sentido descendente por el número de citas que tienen en esa materia, se empiezan a encontrar una serie de núcleos o zonas que detectan cuales son las revistas más importantes en esa materia.

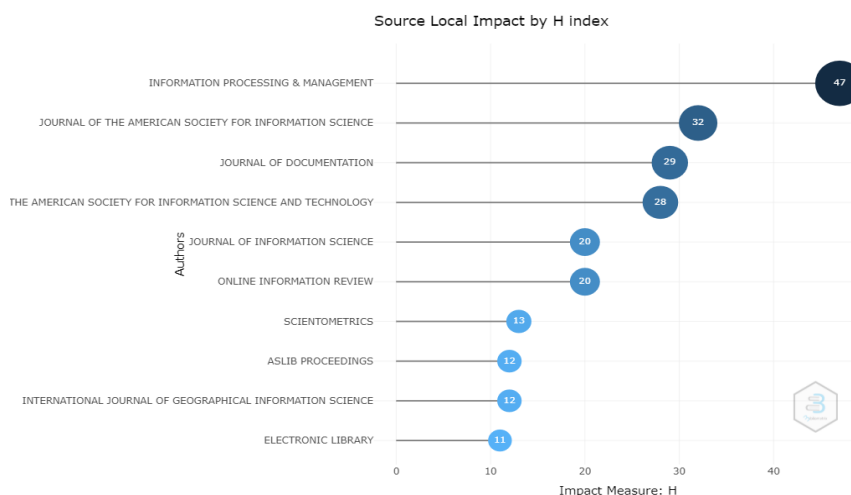
Bradford habla de 3 núcleos y dice que aquellas que están en la zona o núcleo 1 son las revistas más importantes en una materia.

La figura 5 nos muestra cuatro títulos de revistas que según la Ley de Bradford son las que más se consultan en un centro de información y son las que mejor representan el contenido de los documentos.

De nuevo, Information Processing & Management es la más consultada, seguida de Journal of Documentation, Online Information Review y Electronic Library. Observar que todas ellas están en lo que Bradford denomina el Núcleo o Zona 1 por lo tanto, y teniendo en cuenta su ley, son las más importantes en RI.

### 3.3.4. Revistas con Mayor Índice H

El índice H o H-Index es un indicador que sirve para detectar la importancia de un autor, una revista, una categoría..., en base a la cantidad de citas que recibe.



Revistas	h_index	Total de Citas	Nº de Artículos	Año de Aparición
12TH NATIONAL ONLINE MEETING : PROCEEDINGS 1991	1	4	2	1991
13TH NATIONAL ONLINE MEETING : PROCEEDINGS - 1992	1	2	2	1992
14TH NATIONAL ONLINE MEETING : PROCEEDINGS 1993	1	34	1	1993
16TH INTERNATIONAL CONFERENCE ON SCIENTOMETRICS & INFORMETRICS (ISSI 2017)	1	7	2	2017
17TH INTERNATIONAL CONFERENCE ON SCIENTOMETRICS & INFORMETRICS (ISSI2019), VOL I	1	1	1	2019

19TH ANNUAL NATIONAL ONLINE MEETING, PROCEEDINGS	1	6	3	1998
19TH ANNUAL NATIONAL ONLINE MEETING, PROCEEDINGS-1998	1	6	3	1998
2003 JOINT CONFERENCE ON DIGITAL LIBRARIES, PROCEEDINGS	1	2	1	2003
2013 INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND APPLICATIONS (ICISA 2013)	1	1	1	2013
2013 SECOND IIAI INTERNATIONAL CONFERENCE ON ADVANCED APPLIED INFORMATICS (IIAI-AAI 2013)	1	3	2	2013
2014 IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES (JC DL)	3	20	5	2014

Figura 6. Revistas con mayor índice H.

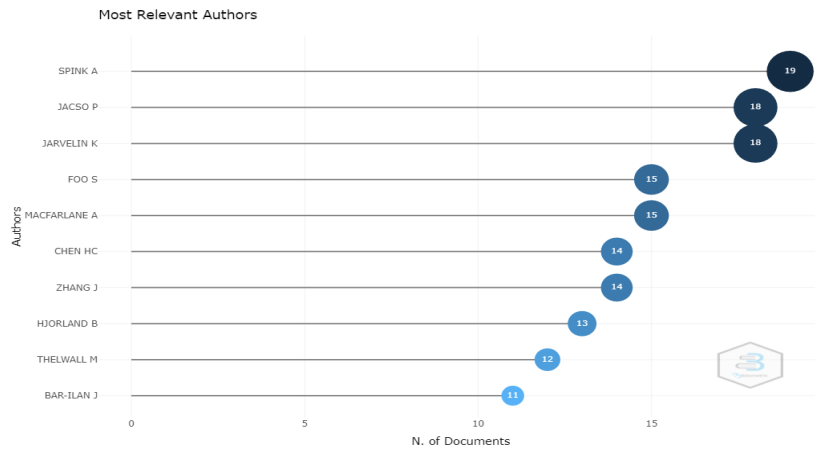
El índice H o H-Index es un indicador que desarrolló Hirsh y sirve para detectar la importancia de un autor, una revista, una categoría.. en base a la cantidad de citas que recibe.

Es un índice fácil de calcular, que favorece a los investigadores más antiguos y perjudica a los investigadores más jóvenes que tienen menos producción científica y por lo tanto menos posibilidad de ser citados.

El Índice H se calcula en base al cruzar el orden de las revistas más citadas en sentido descendente. Según esto, las mejores revistas son las que se observan en la figura 8 de este trabajo, destacando de nuevo Information Processing & Management y seguida de Journal of the American Society for Information Science.

### 3.4. Dimensión de Autores

#### 3.4.1 Autores más relevantes (Producción)



AUTORES	ARTICULOS
SPINK A	19
JACSO P	18
JARVELIN K	18
FOO S	15
MACFARLANE A	15
CHEN HC	14
ZHANG J	14
HJORLAND B	13
THELWALL M	12
BAR-ILAN J	11

Figura 7. Producción por autores

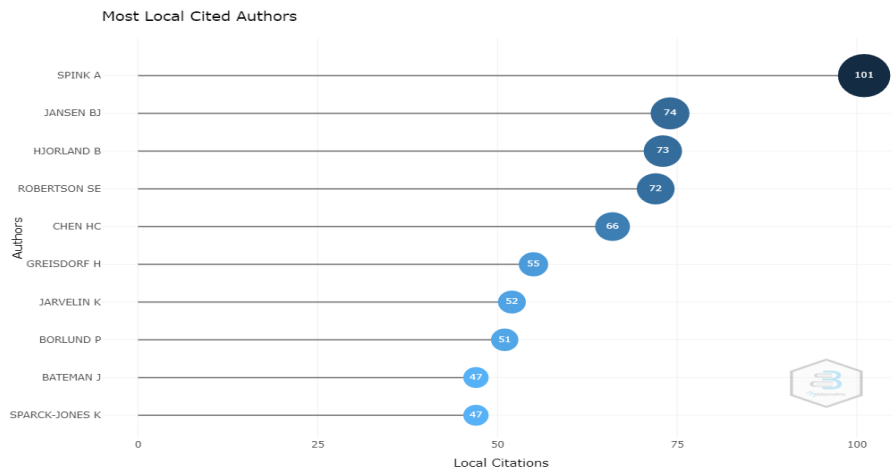
La siguiente gran perspectiva analizada es la dimensión de los autores. Conocer quiénes son los autores más importantes en la RI nos ayudará a investigar sobre sus trabajos, su trayectoria, y serán una fuente de referencia importante para cualquier investigador interesado en la Recuperación de la Información.

Pues bien, analizando esta dimensión se conoce la producción de los autores más importantes en esta materia. Pero además se muestran cuáles son los autores más citados localmente, es decir, los que mayor número de citas han recibido de los documentos descargados entre ellos.



En primer lugar aparece Amanda Spink, con 101 citas locales seguida de Jansen con 74 citas. Los siguientes autores de entre los 10 más citados entre ellos tienen pocas diferencias en el número de citas siendo 51 las citas de Borlund, P que ocupa el décimo lugar.

### 3.4.1.1 Autores más citados localmente



Autores	Citas Locales
SPINK A	101
JANSEN BJ	74
HJORLAND B	73
ROBERTSON SE	72
CHEN HC	66
GREISDORF H	55
JARVELIN K	52
BORLUND P	51

Figura 8. Autores Localmente Citados

En cuanto a los autores más citados en el tiempo, si observamos el gráfico de la figura 11 apreciamos una línea roja. Esta representa la línea temporal, es decir, durante cuánto tiempo ha sido productivo ese autor.

Se representa la producción y los artículos más relevantes o más citados que han escrito estos autores. Cada círculo que observamos es la producción que ha tenido ese autor o autora. Es mayor o menor en función del número de artículos que tiene en cada año. Además el gráfico muestra cuales son los artículos más citados, es decir, muestra la producción y el impacto que tiene.

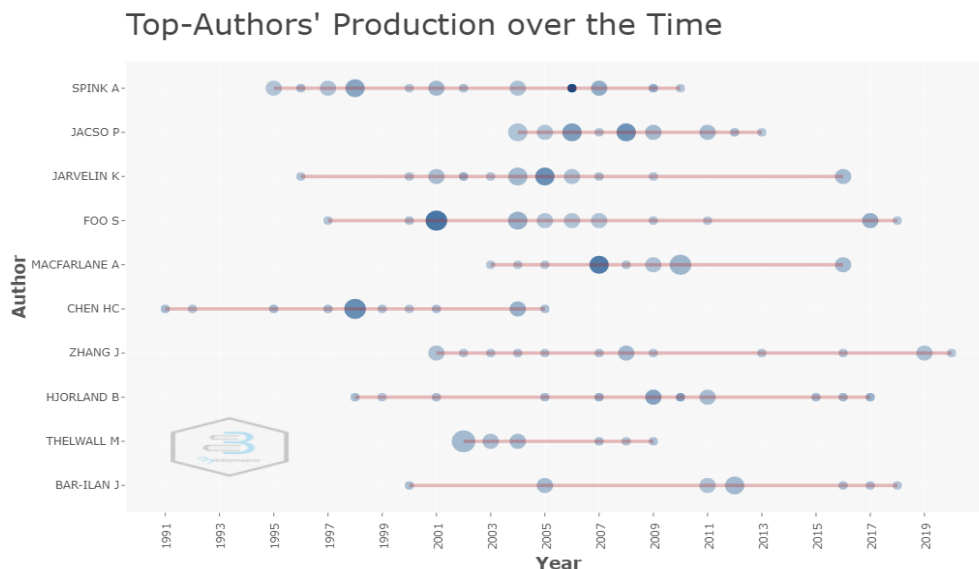
El primer puesto lo ocupa Amanda Spink. Esta autora empieza a producir con citas en el año 1995. En este año ha producido 2 artículos y una media de 0.15 citas por año. Desde el año 1995 hasta el año 2020 esta autora ha tenido unas 22 o 23 citas.

En el año 1996 tiene solamente 1 documento pero con una media de 2.04 citas. Por lo tanto este artículo es más importante porque es más citado.

El segundo puesto lo ocupa Jacso, P. En el año 2004 con un total de 3 artículos tiene una media de 0.44 citas. Pero en el año 2008 con 3 artículos publicados tiene una media de 10.57 citas, por lo tanto este artículo es muy utilizado por otros autores.

### 3.4.1.2 Autores más citados localmente

En esta figura se muestra la producción de una serie de autores. La línea roja representa la línea temporal, es decir, durante cuánto tiempo ha sido productivo ese autor.



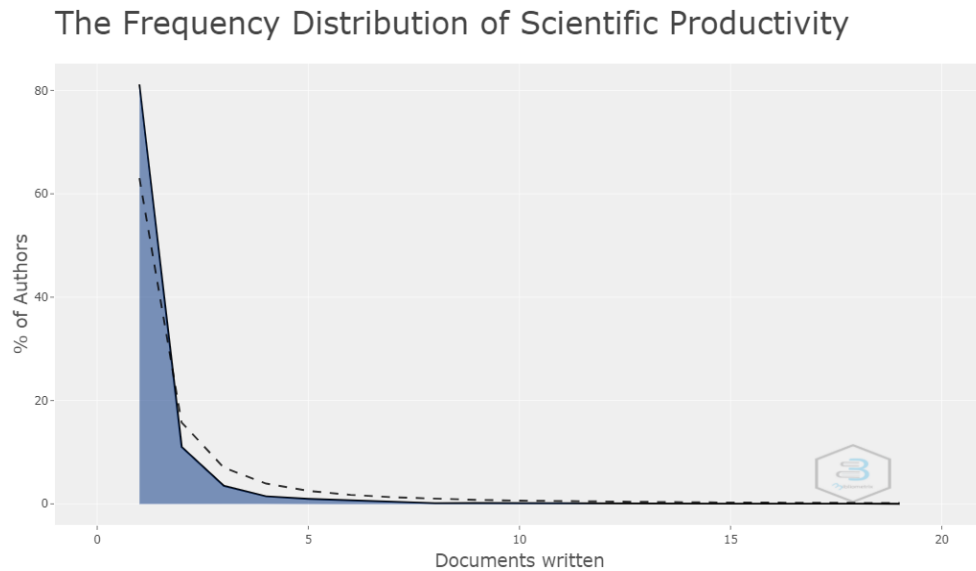
Autor	Año	Título del Art.	Revista	Total Citas	Total Citas por Año
SPINK A	2009	PATTERNS OF QUERY REFORMULATION DURING WEB SEARCHING	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	66	5,077
SPINK A	2007	DEFINING A SESSION ON WEB SEARCH ENGINES	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	68	4,533
SPINK A	2007	COGNITIVE SHIFTS RELATED TO INTERACTIVE INFORMATION RETRIEVAL	ONLINE INFORMATION REVIEW	8	0,533
SPINK A	2006	HOW ARE WE SEARCHING THE WORLD WIDE WEB? A COMPARISON OF NINE SEARCH ENGINE TRANSACTION LOGS	INFORMATION PROCESSING & MANAGEMENT	394	24,625
SPINK A	2004	ELICITATION PURPOSES AND TASKS DURING MEDIATED INFORMATION SEARCH	JOURNAL OF DOCUMENTATION	3	0,167
SPINK A	2004	SEARCHING FOR PEOPLE ON WEB SEARCH ENGINES	JOURNAL OF DOCUMENTATION	13	0,722
SPINK A	2002	ISSUES OF CONTEXT IN INFORMATION RETRIEVAL (IR): AN INTRODUCTION TO THE SPECIAL ISSUE	INFORMATION PROCESSING & MANAGEMENT	31	1,55
SPINK A	2001	MEDIAN MEASURE: AN APPROACH TO IR SYSTEMS EVALUATION	INFORMATION PROCESSING & MANAGEMENT	16	0,762
SPINK A	2001	REGIONS AND LEVELS: MEASURING AND MAPPING USERS' RELEVANCE JUDGMENTS	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	39	1,857
SPINK A	2000	RECENT RELEVANCE RESEARCH: IMPLICATIONS FOR INFORMATION PROFESSIONALS	ONLINE INFORMATION REVIEW	7	0,318

Figura 9. Autores más citados en el tiempo

En la figura anterior, se muestra cuáles son los autores más citados en el tiempo y durante cuanto tiempo cada autor ha sido muy productivo. La línea roja muestra el tiempo durante el que cada autor ha sido productivo, y los círculos que se observan, indican el momento de mayor productividad de cada uno de ellos. Cuanto más grande el círculo y más oscuro, mayor ha sido la productividad del sujeto.

### 3.4.1.3 Productividad según la Ley de Lotka

La Ley de Lotka describe la frecuencia de publicaciones por autor en un determinado campo.



Artículos	Número de Autores	Proporción de Autores
1	3397	0,811
2	462	0,11
3	146	0,035
4	61	0,015
5	41	0,01
6	28	0,007
7	17	0,004
8	6	0,001
9	7	0,002
10	6	0,001

Figura 10. Ley de Lotka

En cuanto a la perspectiva de la productividad según la Ley de Lotka definir que esta Ley de Lotka describe la frecuencia de publicaciones por autor en un determinado campo.

Hace una aproximación inversa al cuadrado, donde el número de autores que publica una determinada cantidad de artículos se fija de acuerdo con un ratio en base a la cantidad de artículos que se publican.

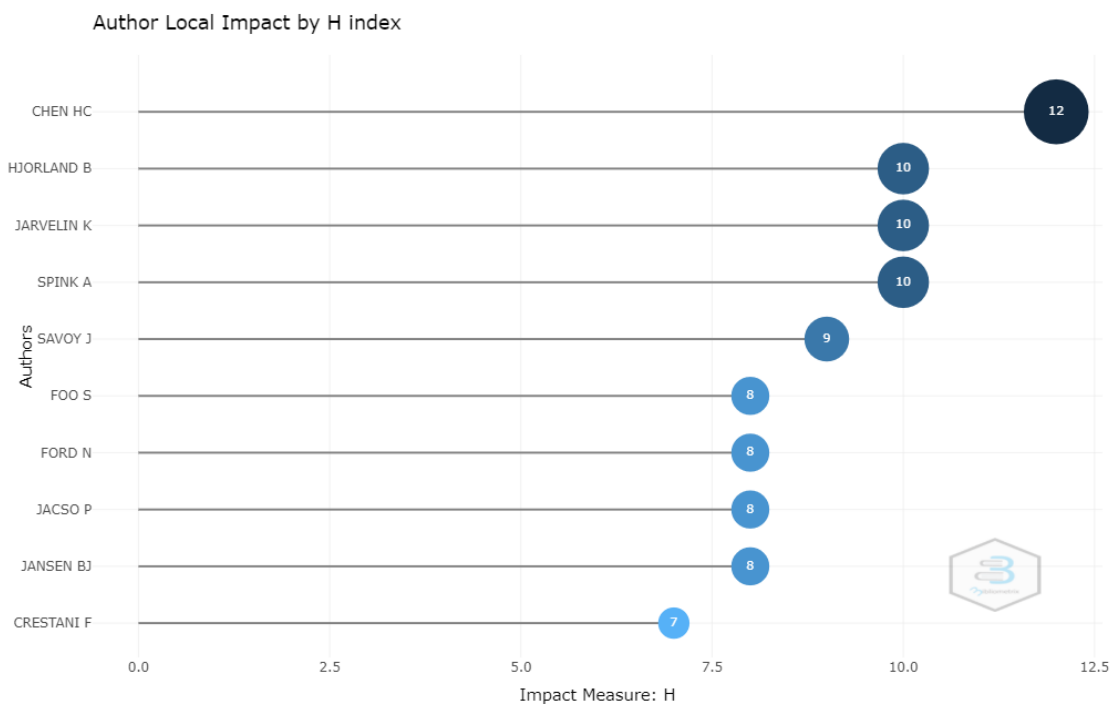
En la medida en que aumenta el número de artículos publicados, los autores que publican muchas publicaciones se reducen.

Según la ley de Lotka, cuantos más artículos hay, el número de autores que producen muchos artículos disminuye.

La Ley de Lotka identifica cuales son los autores principales a partir de los documentos a partir de los cuales se considera que estos autores representan la totalidad de artículos producidos.

El gráfico representa a los mejores trabajos, los más representativos en esta área.

### 3.4.1.4 Autores con mayor impacto (Indice H)



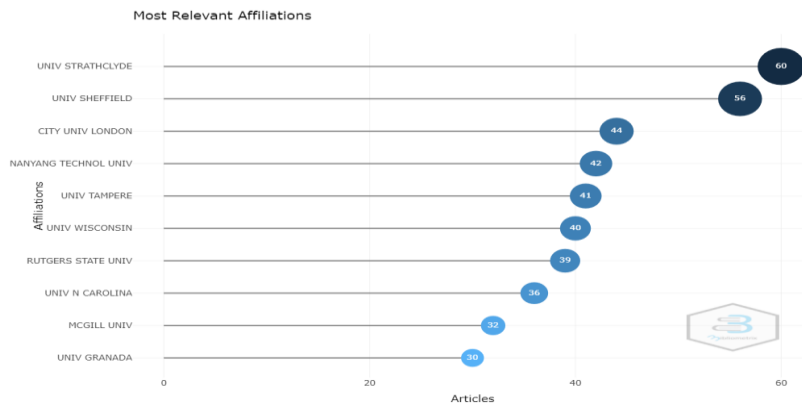
Autor	Índice H	Total de Citas	Num. de Publicac	Año Comienzo Publicar
CHEN HC	12	634	13	1991
HJORLAND B	10	431	13	1998
JARVELIN K	10	441	18	1996
SPINK A	10	870	18	1995
SAVOY J	9	212	11	1992
FOO S	8	471	12	1997
FORD N	8	349	8	2002
JACSO P	8	371	15	2004
JANSEN BJ	8	768	8	2004
CRESTANI F	7	177	9	2000

Figura 11. Autores con mayor impacto (Índice H)

La siguiente perspectiva muestra a los autores con mayor impacto o Índice H.

El autor con mayor impacto o Índice H es Chen, HC con un total de 634 citas. Las publicaciones de este autor comienzan a tener importancia en el año 1991, por lo que al ser tan antiguo es normal que tenga más citas ya que tiene más posibilidades de tener un mayor Índice H.

### 3.4.1.5 Instituciones donde publican los autores más productivos



Institución	Artículos
UNIV STRATHCLYDE	60
UNIV SHEFFIELD	56
CITY UNIV LONDON	44
NANYANG TECHNOL UNIV	42
UNIV TAMPERE	41
UNIV WISCONSIN	40
RUTGERS STATE UNIV	39
UNIV N CAROLINA	36
MCGILL UNIV	32
UNIV GRANADA	30

Figura 12. Instituciones más importantes en las que publican los autores

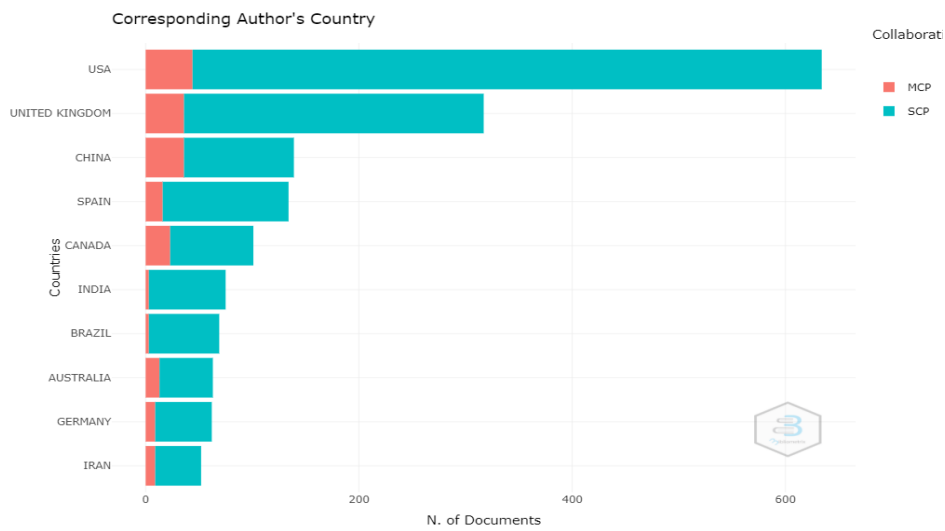
Biblioshiny también muestra una perspectiva muy interesante. Mediante gráficos podemos ver cuáles son las instituciones más importantes en las que publican los autores.

Las instituciones que más artículos publican en RI son Universidad de Strathclyde con 60 artículos, seguida por la universidad de Sheffield con 56 artículos. En quinto lugar está la Universidad de Tampere en Finlandia. Señalar que esta universidad el principal autor es Karl Jarveli que es uno de los editores de la revista Information Processing & Management.

Mencionar a la Universidad de Granada que está en el décimo puesto en cuanto a publicación de artículos relacionados con la RI, un lugar nada despreciable.

### 3.4.1.6 Autoría de un Artículo

Con esta perspectiva se determina quién se atribuye el mérito de un artículo.



País	Artículos	Frecuencia	Sin colaboración	Múltiple colaboración	Ratio
USA	634	0,268758	590	44	0,0694
UNITED KINGDOM	317	0,134379	281	36	0,1136
CHINA	139	0,058923	103	36	0,259
SPAIN	134	0,056804	118	16	0,1194
CANADA	101	0,042815	78	23	0,2277
INDIA	75	0,031793	72	3	0,04
BRAZIL	69	0,02925	66	3	0,0435
AUSTRALIA	63	0,026706	50	13	0,2063
GERMANY	62	0,026282	53	9	0,1452
IRAN	52	0,022043	43	9	0,1731

Figura 13. Países a los que pertenecen los autores que más publican en RI

En cuanto a la Autoría de los artículos uno de los problemas que se plantean a veces con relación a la producción científica y saber quién es el autor de un determinado artículo y una determinada institución es saber quién se atribuye el mérito de un artículo.

En el campo de las Ciencias Sociales el primer autor se considera que es el responsable.

En otras ciencias como las Matemáticas o Medicina es el último.

También depende del número de autores que participan en la redacción del artículo.



Para poder discernir cual es de verdad el autor de un artículo se ha acordado que el autor que aparece como corresponding author sea el autor principal.

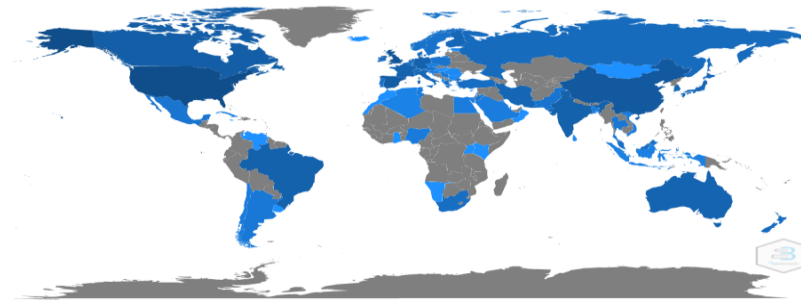
En el caso de la figura 13 se observa que el país que más produce sobre RI en el mundo es Estados Unidos, pero además colabora con otros países pero el autor principal es de Estados Unidos.

Reino Unido tiene 50 trabajos y 276 documentos donde es la figura principal.

En el caso de España, la proporción de documentos que tiene colaborando con otros países y donde se lidere es poco. La mayoría de los trabajos los hacen los autores sin colaborar con otros países. No se colabora y esto es perjudicial porque cuanto más se colabora más visibilidad se alcanza y más impacto tiene la publicación.

### 3.4.1.7 Países con más producción científica

Country Scientific Production



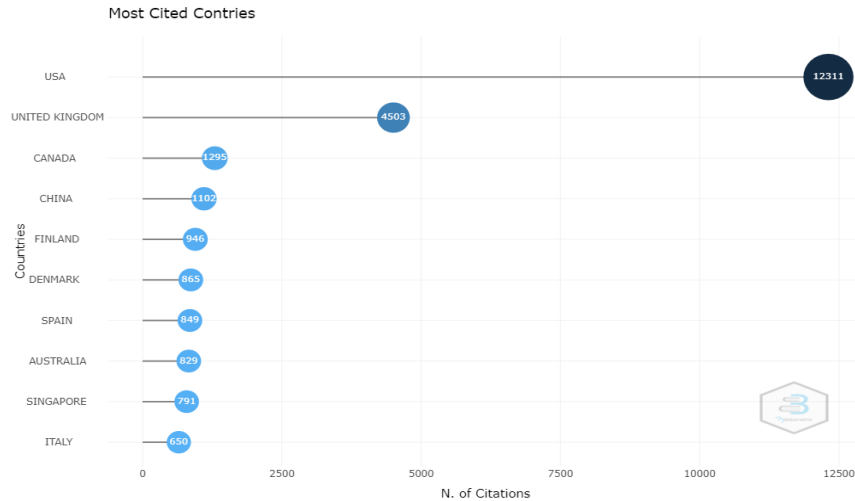
País	Frecuencia
USA	918
UK	477
CHINA	271
SPAIN	187
CANADA	166
INDIA	124
BRAZIL	123
GERMANY	103
FRANCE	99
AUSTRALIA	90

Figura 14. Países con más producción científica en RI

De entre los 10 primeros países que aparecen en la tabla el país que más produce es Estados Unidos con 918 artículos. Le siguen Reino Unido y China y en cuarto lugar está España con 187 artículos acerca de RI.

### 3.4.1.8 Países más citados

No todo lo que se produce se utiliza o tiene calidad.



País	Total Citas	Promedio Citas por Artículo
USA	12311	19,42
UNITED KINGDOM	4503	14,21
CANADA	1295	12,82
CHINA	1102	7,93
FINLAND	946	22,00
DENMARK	865	27,90
SPAIN	849	6,34
AUSTRALIA	829	13,16
SINGAPORE	791	19,29
ITALY	650	16,67

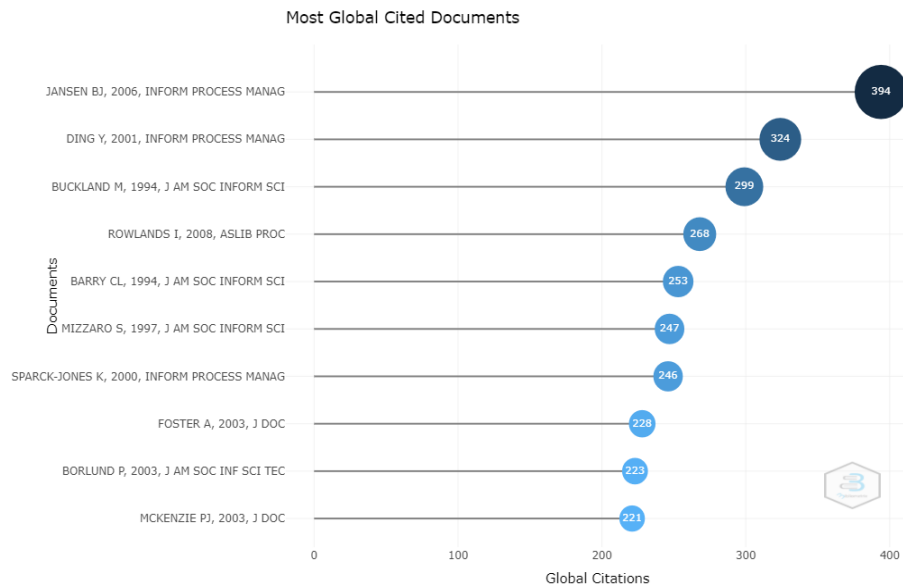
Figura 15. Países más citados

En cuanto a los países más citados cabe destacar que no todo lo que se produce se utiliza o tiene mucha calidad. Muchas de las cosas que se publican no se usan mucho, es decir, no todo tiene calidad. No todo lo que publican algunos países se usa para investigar. Como ejemplo, y comparando las figuras 16 y 17, en lo relacionado entre producción científica y países más citados surge Canadá. Este país aparece en el quinto puesto en lo referente

a producción científica, sin embargo sube hasta el puesto tercero en lo relacionado a países más citados. Este tercer lugar lo ocupa China en el gráfico de país con mayor producción científica pero no todo lo que produce es de calidad y por esto Canadá lo adelanta como país más citado.

## 3.5 Documentos

### 3.5.1 Trabajos más citados a nivel global



36

Documento	Total de Citas	Total Citas por año	Total Citas Normalizadas
JANSEN BJ, 2006, INFORM PROCESS MANAG	394	24,63	18,56
DING Y, 2001, INFORM PROCESS MANAG	324	15,43	18,08
BUCKLAND M, 1994, J AM SOC INFORM SCI	299	10,68	7,57
ROWLANDS I, 2008, ASLIB PROC	268	19,14	16,77
BARRY CL, 1994, J AM SOC INFORM SCI	253	9,04	6,41
MIZZARO S, 1997, J AM SOC INFORM SCI	247	9,88	9,98
SPARCK-JONES K, 2000, INFORM PROCESS MANAG	246	11,18	12,48
FOSTER A, 2003, J DOC	228	12,00	8,49
BORLUND P, 2003, J AM SOC INF SCI TEC	223	11,74	8,31
MCKENZIE PJ, 2003, J DOC	221	11,6316	8,23

Fig. 16. Documentos más citados

Una de estas perspectivas son los trabajos más citados a nivel global. El gráfico 16 muestra los 10 trabajos más importantes a nivel global que se extraen de la base de datos WOS acerca de RI, es decir, los trabajos que cualquier investigador debería consultar para comenzar a trabajar sobre este tema.

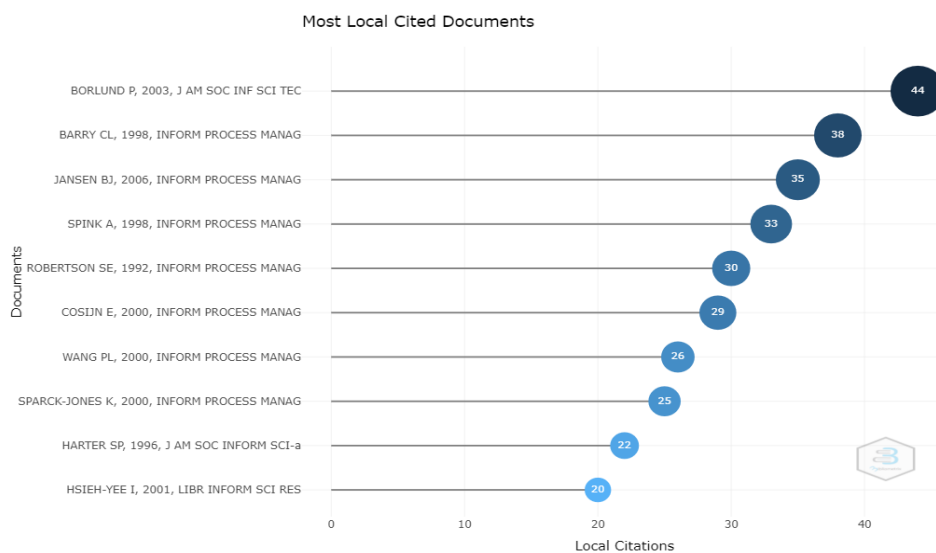
Hay que tener en cuenta que estos son trabajos que pueden citar investigadores de cualquier área y no solo de Recuperación de Información. Por esto, se dice que son los trabajos más importantes a nivel global.

El primer trabajo es “How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Lo publica Bernard J. Jansen en el año 2006, en la revista *Information Processing & Management*. Ha tenido un total de citas de 394 y la media de citas por año es de 24.650.

Le sigue el artículo de Ying Ding “Bibliometric cartography of information retrieval research by using co-word analysis publicado en la revista *Information Processing & Management* en el año 2001. Este artículo tiene un total de citas de 324 y la media de citas por año es de 15.486.

El resto de artículos de entre los 10 más citados tienen entre las 299 y las 221 citas del último artículo que destaca. Este es el artículo de Pamela J. McKenzie “A model of information practices in accounts of everyday-life information seeking”. Se publicó en el año 2003 en la revista *Journal of Documentation*.

### 3.5.1.1. Documentos más citados a nivel local



DOI	Año	Citas Locales	Citas Globales	LC/GC Ratio (%)
10.1002/asi.10286	2003	44	223	19,73
10.1016/S0306-4573(97)00078-2	1998	38	159	23,90
10.1016/j.ipm.2004.10.007	2006	35	394	8,88
10.1016/S0306-4573(98)00025-9	1998	33	151	21,85
10.1016/0306-4573(92)90004-J	1992	30	106	28,30
10.1016/S0306-4573(99)00072-2	2000	29	109	26,61
10.1016/S0306-4573(99)00059-X	2000	26	158	16,46
10.1016/S0306-4573(00)00015-7	2000	25	246	10,16
	1996	22	98	22,45
10.1016/S0740-8188(01)00069-X	2001	20	89	22,47

Figura 17. Autores más citados a Nivel Local

Según los autores que tratan el tema de RI este gráfico contesta a la pregunta de ¿quiénes son los autores más relevantes, más citados en RI?

A diferencia del gráfico anterior, en este caso se muestran los autores más relevantes o más citados en RI pero solamente entre ellos, esto es, a nivel local, y no de otras áreas como sucedía en el gráfico de los Documentos más citados a nivel global.

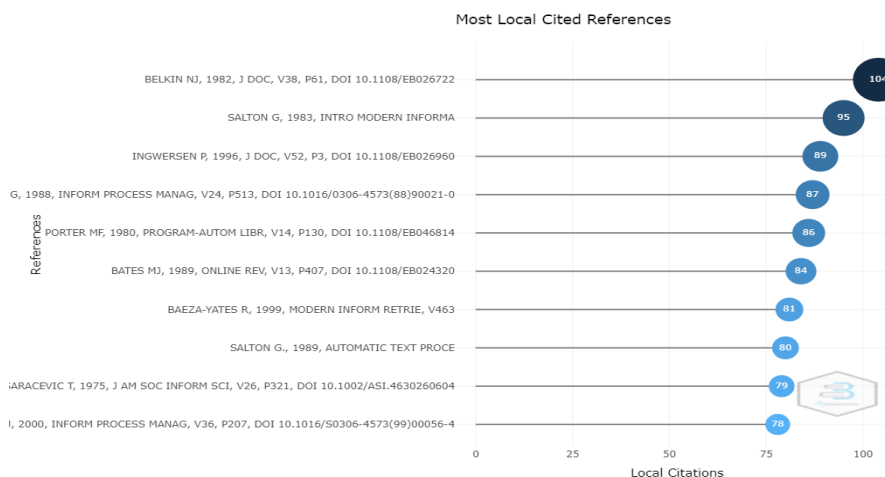
Como observamos en el gráfico el trabajo más citado es el publicado en el año 2003 por Pía Borlund en la revista Journal of the American Society for Information Science and Technology y titulado “The concept of relevance in IR”.

Este trabajo tiene un total de citas locales de 44 y un total de citas globales de 223. La ratio entre las citas locales y globales es de 19.73%, lo que lo sitúa en el número uno de entre los trabajos publicados para la Recuperación de la Información.

En segunda posición se sitúa el trabajo de Carol L. Barry “Users’ criteria for relevance evaluation: A cross-situational comparison. Este artículo fue publicado en el año 1998 en la revista Information Processing & Management y tiene un total de 38 citas locales y 159 citas globales, siendo la ratio entre las citas locales y las globales del 23.90%.

El número de citas locales continúa descendiendo poco hasta llegar al décimo trabajo de entre los diez mejores. Este es el artículo publicado por Ingrid Hsieh-Yee en el año 2001 “Research on Web search behavior”, en la revista Library & Information Science Research. El número de citas es de 20 a nivel local y 89 a nivel global. La ratio entre ambos niveles de citas es de 22.47%.

### 3.5.1.2 Documentos fuente más citados en el área de RI



Referencias	Citas
BELKIN NJ, 1982, J DOC, V38, P61, DOI 10.1108/EB026722	104
SALTON G, 1983, INTRO MODERN INFORMA	95
INGWERSEN P, 1996, J DOC, V52, P3, DOI 10.1108/EB026960	89
SALTON G, 1988, INFORM PROCESS MANAG, V24, P513, DOI 10.1016/0306-4573(88)90021-0	87
PORTER MF, 1980, PROGRAM-AUTOM LIBR, V14, P130, DOI 10.1108/EB046814	86
BATES MJ, 1989, ONLINE REV, V13, P407, DOI 10.1108/EB024320	84
BAEZA-YATES R, 1999, MODERN INFORM RETRIE, V463	81
SALTON G., 1989, AUTOMATIC TEXT PROCE	80
SARACEVIC T, 1975, J AM SOC INFORM SCI, V26, P321, DOI 10.1002/ASI.4630260604	79
JANSSEN BJ, 2000, INFORM PROCESS MANAG, V36, P207, DOI 10.1016/S0306-4573(99)00056-4	78

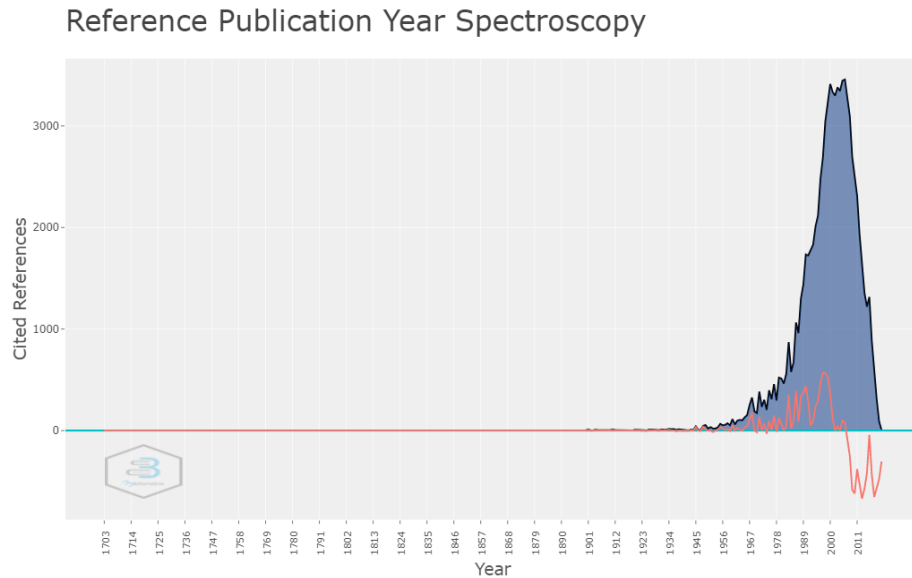
Figura 18. Referencias a nivel local más citadas

De estos datos se deducen cuáles serían los 10 artículos que de la materia de Recuperación de la Información debería consultar alguien que quiera investigar sobre esto, ya que son

los artículos más citados entre ellos mismos y nadie mejor que los propios autores para decir de qué tratan sus escritos.

Notamos la diferencia de puesto que ocupan los artículos si los analizamos a nivel local o a nivel global.

### 3.5.1.3 Espectroscopia de las Referencias



AÑO	REFERENCIA	CITAS LOCALES
2020	TAHERY S 2020 INFORM SYST	2
2020	ABDELNABY M. 2020 ALEXANDRIA ENG J	1
2020	ADAMS J. 2020 WHAT ROLE DOES COLLA	1
2020	ANONYMOUS 2020 COVID 19 ESPECIALIDA	1
2020	ARROYUELO D 2020 INFORM SYST	1
2020	AVILA BARRIENTOS EDER 2020 INVESTIG BIBLIOTECOL	1
2020	BAO YY 2020 CHEM MATER	1

Figura 19. Espectroscopia

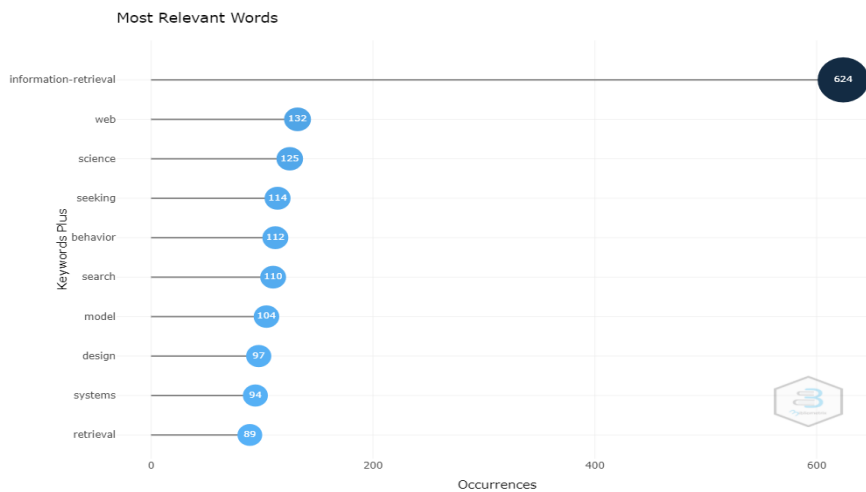


La perspectiva de las Espectroscopia nos muestra de qué años se utiliza el conocimiento anterior. Se están citando trabajos del año 1703, 1714, 1802 etc. pero es a partir del año 1949 y sobre todo documentos que hay hasta el año 2006 cuando se comienzan a utilizar referencias con mayor asiduidad.

Estos no son documentos publicados hasta el año 2006 y a partir de aquí desciende la cantidad. Se refiere a documentos que más se utilizan de años anteriores y son desde el año 1949 y hasta el año 2006.

La línea roja es una línea de tendencia que representa el número de referencias que se van utilizando en ese tiempo. En el año 2007 la proporción es negativa y así continúa hasta 2011.

### 3.5.1.4 Palabras más frecuentes



PALABRAS	FRECUENCIA
information retrieval	1555
internet	143
search engines	115
digital libraries	79
user studies	72
classification	60
databases	60

evaluation	59
indexing	57
worldwide web	55

Figura 20. Palabras más frecuentes según Author's Keywords

En cuanto a las palabras más frecuentes utilizadas se observa en la figura 20 que el término information retrieval aparece 1555 veces. Esto es lógico porque se ha buscado en la base de datos ese término. Lo más relevante lo indica los siguientes términos porque la frecuencia que aparece a partir de Inforamation Retrieval es lo más esperado.

### 3.5.1.5 WordCloud

El número de palabras que se representan es de 100 porque hasta aquí el ojo humano las distingue perfectamente.



TÉRMINOS	FRECUENCIA
information-retrieval	624
web	132
science	125
seeking	114
behavior	112
search	110
model	104

design	97
systems	94
retrieval	89

Figura. 21. Nube de Palabras

Las palabras más utilizadas, también se representan con la figura de la llamada nube de palabras. Observar que aparece en el centro behavior, web science..., pero hay que destacar que no aparece Information Retrieval porque se utiliza la Ley de Zipf: aquellos términos que aparecen muchas veces no son representativos. Es decir, aparece tantas veces que no es representativo en este caso de todas las relaciones que se muestran.

### 3.6 ESTRUCTURAS DEL CONOCIMIENTO

Cuando se habla de estructuras del conocimiento se hace referencia a la representación gráfica de ese conocimiento. El vocablo en inglés que se utiliza para referirse a esto es Big Picture, lo que quiere decir hacer el gran gráfico. Esto ha sido algo que han perseguido todos los investigadores con el objetivo de poder ver de una forma gráfica las estructuras de cualquier área de conocimiento, es decir, que el usuario vea de manera gráfica lo que busca y no solamente mediante registros.

A este tipo de representaciones se les conoce como Science mapping que intenta crear representaciones mediante las conexiones intelectuales, entre los términos que utilizan los autores porque parten de la idea de que cuando un término aparece con otro es porque están relacionadas con él, y si este aparece con otro pasa lo mismo, están interrelacionados.

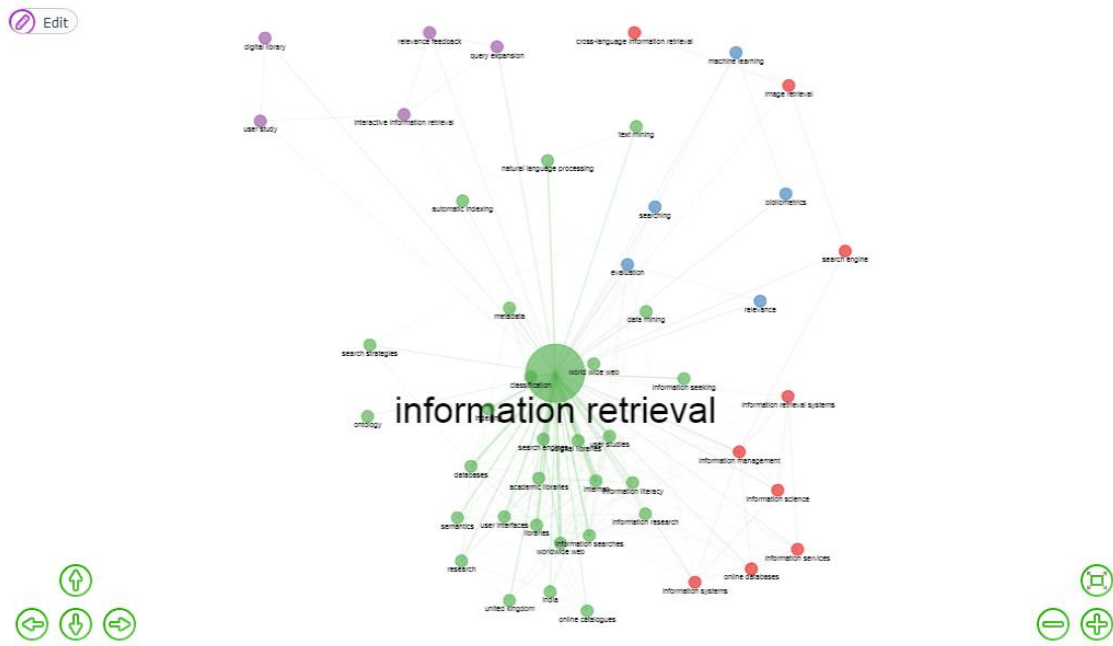
El objetivo es hacer representaciones que pueden mostrar de una forma dinámica o evolutiva como va cambiando el conocimiento a lo largo del tiempo.

Intenta mostrar la estructura y los aspectos dinámicos o evolutivos de la investigación científica y todo esto lo hace a través de la llamada Estructura del Conocimiento.

La comunidad científica pretende tener una visión completa de cuáles son las principales estructuras que están relacionadas con cada campo de la investigación. Pueden ser líneas de investigación grandes como por ejemplo la documentación o líneas más concretas como por ejemplo las bibliotecas escolares, o algo tan concreto como la recuperación de la información.

El Science mapping se puede tratar desde un punto de vista estadístico y se pueden mostrar tres tipos de estructura, la estructura conceptual, la estructura intelectual y la estructura social

### 3.6.1 ESTRUCTURA CONCEPTUAL



NODO	CLUSTER	BETWEENNESS	CLOSENESS	PAGERANK
information retrieval	3	934,26	0,021	0,33
image retrieval	1	46,00	0,011	0,01
search engines	3	4,03	0,013	0,05
internet	3	2,83	0,013	0,05
user studies	3	1,22	0,013	0,04
evaluation	2	0,99	0,012	0,02
databases	3	0,77	0,012	0,03
information management	1	0,68	0,012	0,02
worldwide web	3	0,41	0,012	0,03
libraries	3	0,39	0,012	0,02

Figura 22. Red de Co-ocurrencia

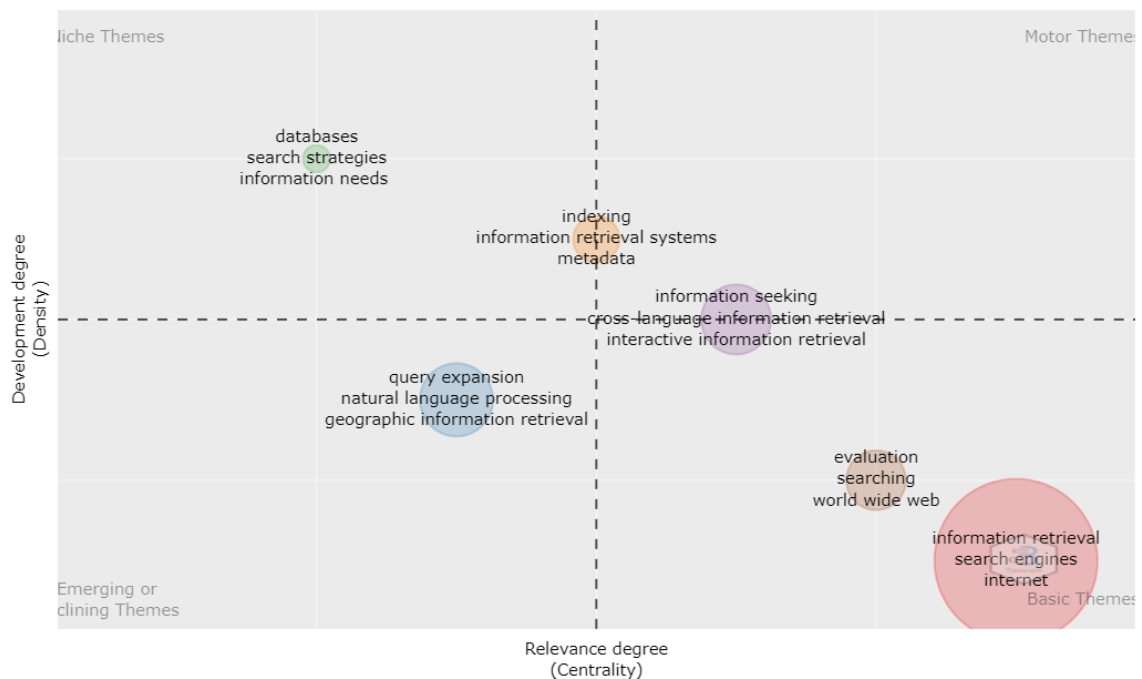
Estructura Conceptual: Hace referencia a qué o sobre qué se habla de la ciencia. Por ejemplo, en el tema de recuperación de información la estructura conceptual muestra cuales son los principales temas de la recuperación de la información y cuáles son las tendencias en esta materia. Esto se hace por medio de las palabras clave y por esto se

habla de una estructura conceptual, porque alguien ha pensado de qué tratan esos artículos y los autores lo han plasmado con esas palabras claves.

Esta estructura representa las relaciones entre los conceptos o palabras entre una serie de publicaciones. Es importante tener un número adecuado de documentos para ver cómo se relacionan las palabras entre ellos.

Las palabras que aparecen juntas en un documento se relacionan entre ellas en forma de red. A ese tipo de relaciones entre los términos se le denomina Red de Co-términos. Esto es así porque aparecen de forma conjunta.

### 3.6.1.1 Estructuras Temáticas



OCCURRENCES	PALABRAS	CLUSTER	CLUSTER_LABEL
1109	information retrieval	1	information retrieval
88	search engines	1	information retrieval
83	internet	1	information retrieval
64	user studies	1	information retrieval
52	digital libraries	1	information retrieval

50	worldwide web	1	information retrieval
40	libraries	1	information retrieval
37	information searches	1	information retrieval
35	classification	1	information retrieval
31	information management	1	information retrieval

Figura 23. Mapa Temático

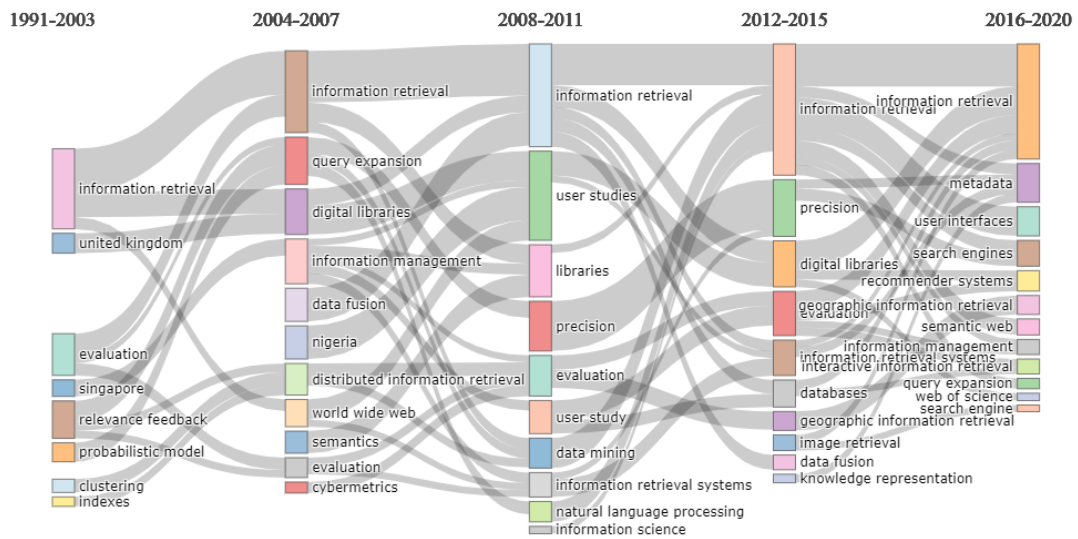


Figura 24. Evolución Temática

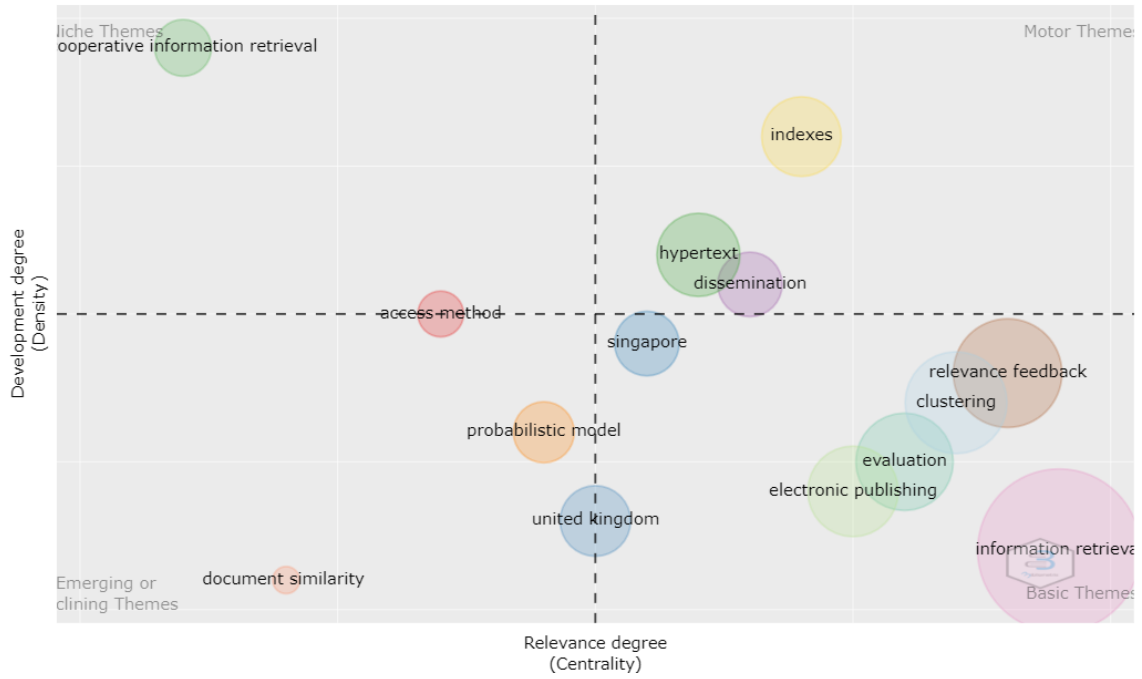


Figura 25. Evolución Temática en los años 1990-2003

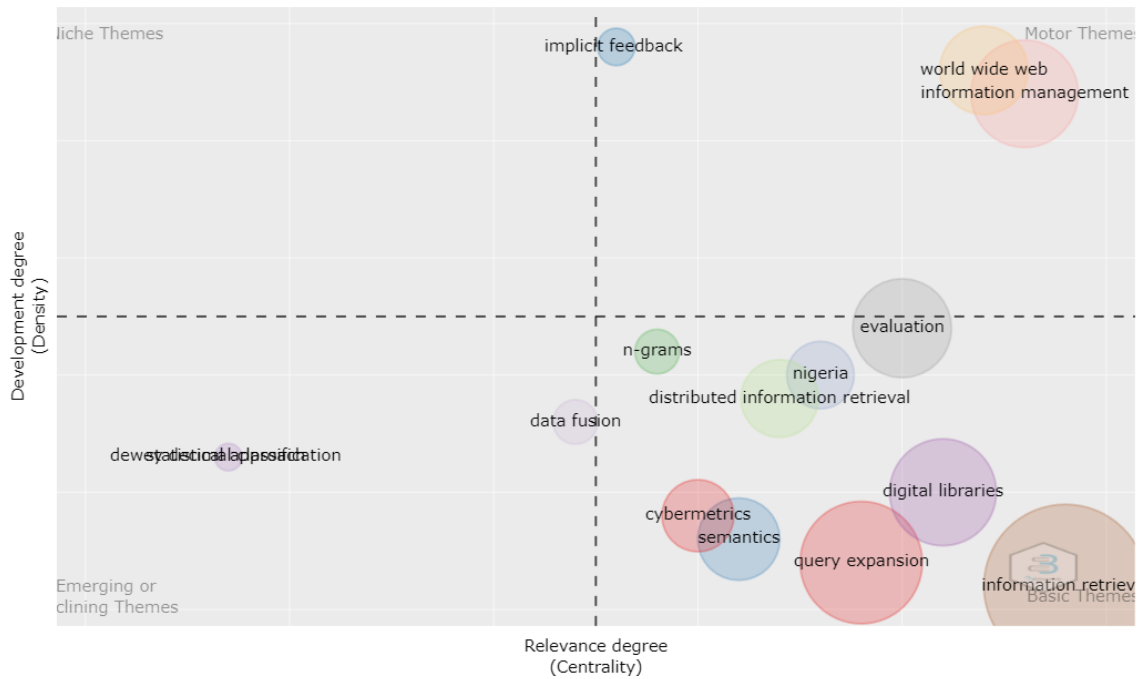


Figura 26. Periodo 2004-2006

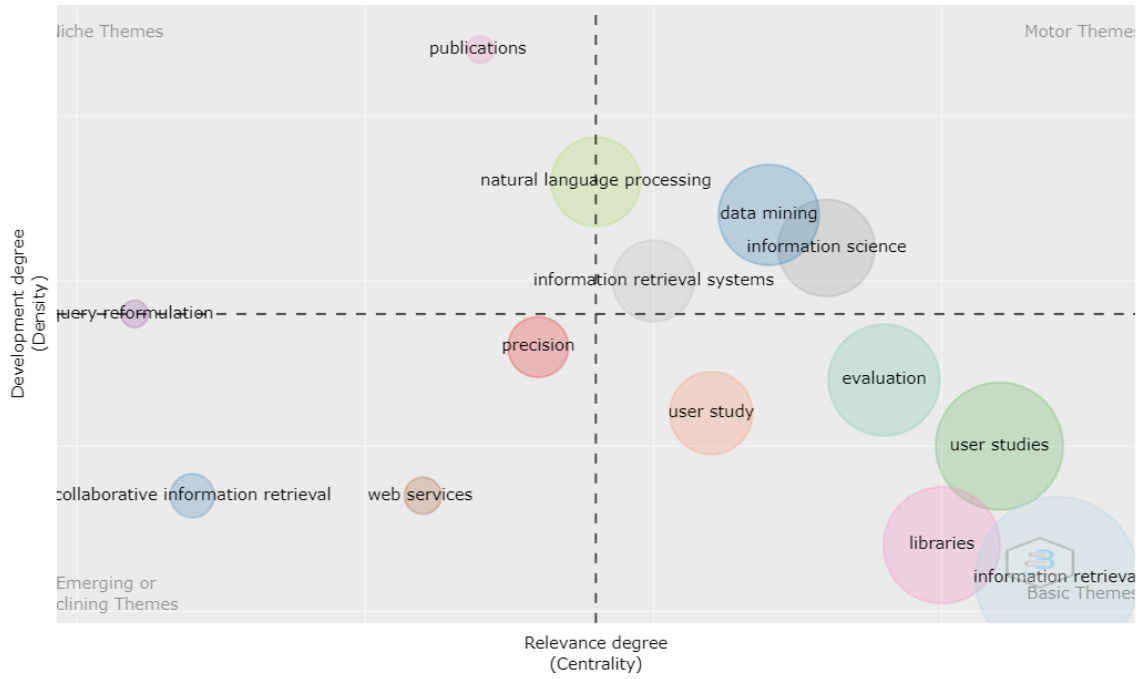


Figura 27. Periodo 2007-2011

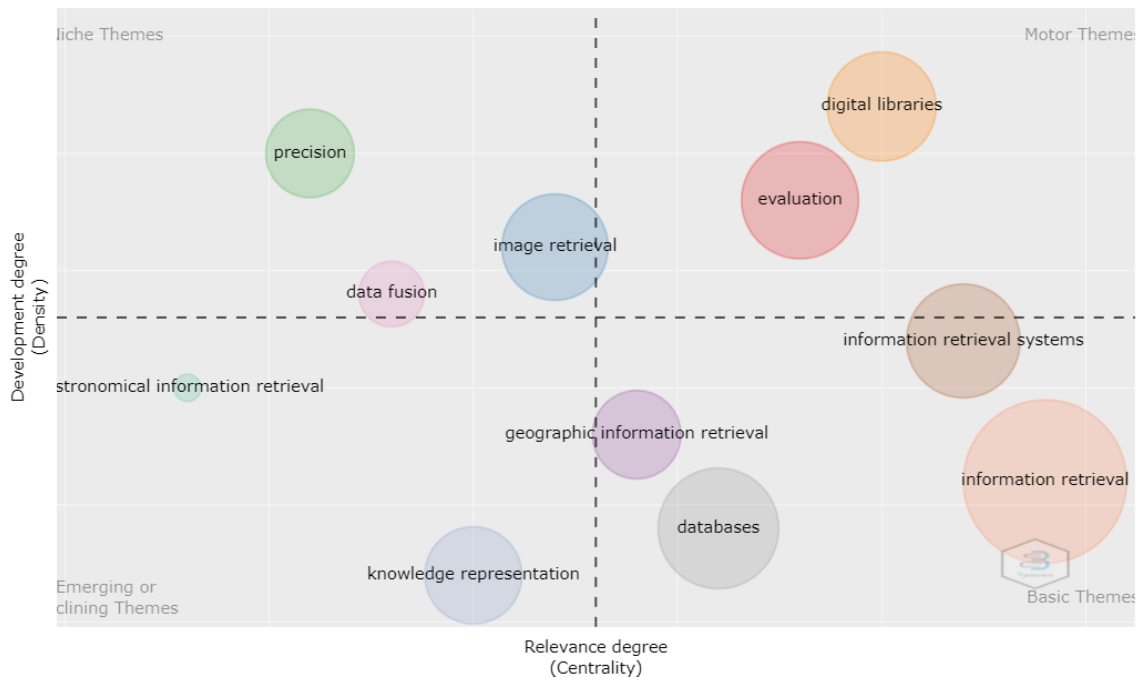


Figura 28. Periodo 2012-2014





Figura 29. Periodo 2015-2020

Esta estructura se puede utilizar para comprender cuales son las principales materias que se tratan en el campo que se está investigando, y cuales son las más importantes y qué temas son los que más se utilizan y a esto se le suele llamar Frente de Investigación.

En este trabajo concretamente, se pueden observar cuales son los términos que más se utilizan del año 1990-2010 y desde 2011 a 2020.

Aplican algoritmos de clustering en las redes. Reducen la dimensión representando los términos agrupados por clúster. Es decir, combina técnicas estadísticas con la visualización reduciendo la dimensión agrupando todos los términos que tratan de lo mismo en un solo clúster y representa solo un término o dos sin son muchos y a todos estos les da el nombre del término que más aparece.

Se representa en cuadrantes y el que está en el cuadrante superior derecho se le llaman temas motor.

Los del cuadrante superior izquierdo son los términos muy desarrollados pero están aislados.

Los temas básicos y transversales son los temas del cuadrante inferior derecho.

Y los términos que están desapareciendo se muestran en el cuadrante inferior izquierdo.

Los más importantes son los temas del cuadrante superior derecho.

La centralidad es la importancia que tiene un tema en un campo. Cuanto más centrado aparece en el gráfico más importante es el tema.

Cuanto más arriba aparece en el gráfico se considera un término muy consolidado.

En la figura 25 se aprecia un mapa temático. En el caso de los términos motor que corresponde al ángulo superior derecho no se observan que haya términos importantes. Por importancia y desarrollo basándonos en el gráfico estaría Sistemas de Recuperación de Información, Metadatos e Indización.

El término Information retrieval es un término básico, que aparece en el cuadrante inferior izquierdo de un gran tamaño por lo que sabemos que engloba una gran cantidad de palabras, y parece que está poco desarrollado.

Este gráfico junto con el gráfico que trata de la evolución temática y la figura de la evolución temática por años nos muestra por donde ha ido la investigación principal por periodos de tiempo. Además, muestra también los temas más importantes pero poco desarrollados y los temas que en su momento fueron importantes, pero ahora la tendencia que tienen es a desaparecer para dejar paso a otros temas que están de actualidad.

### 3.6.1.2 ANÁLISIS FACTORIAL

Tipo de representación en red que muestra los cluster y además los elementos de cada cluster.

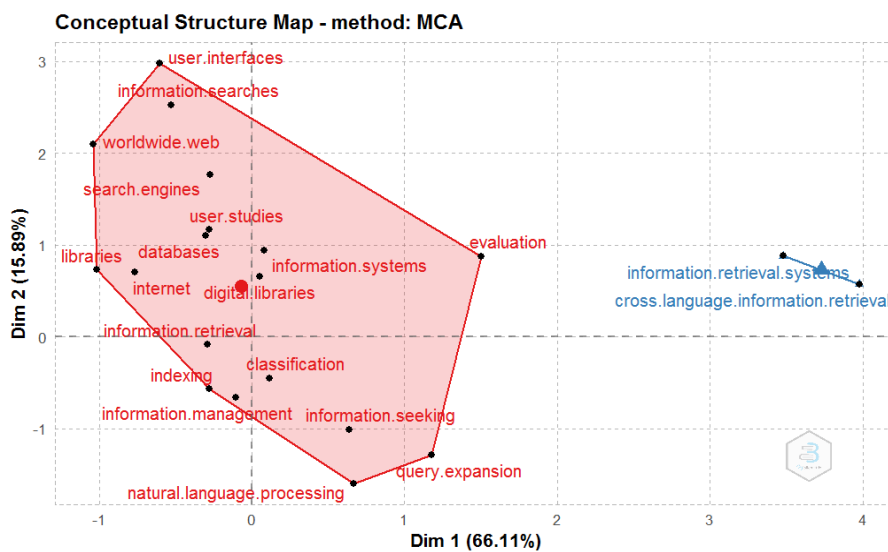


Figura 30. Análisis Factorial

La perspectiva del Análisis Factorial es una técnica estadística que distribuye los elementos en grupos mediante un análisis de términos y agrupándolos en una técnica que se llama Análisis de Componentes Principales, que se llama Análisis Factorial.

Es un tipo de representación en red que muestra los cluster y además los elementos de cada cluster.

La ventaja de este análisis frente al clustering es que en este último un término puede pertenecer a varios clústeres.

El gráfico muestra dos cluster importantes, bibliotecas digitales y recuperación de la información, y los elementos que forman cada cluster.

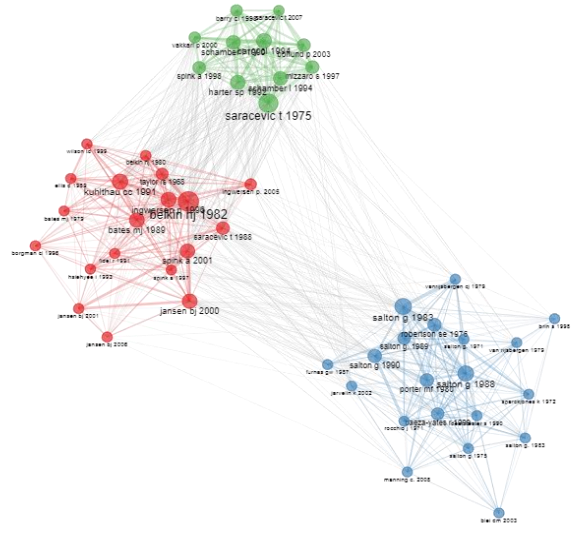
Cuáles son los documentos más importantes de esta materia que se está investigando.

Cuáles son los más citados y cuáles son las palabras por cluster.

### 3.7 ESTRUCTURA INTELECTUAL

Se refiere a cómo los trabajos de unos autores influyen en los trabajos de otros, es decir, como los trabajos de unos autores que ya están escritos son utilizados por otros autores para crear conocimiento. En definitiva, se utilizan las citas

#### 3.7.1 COCITACION



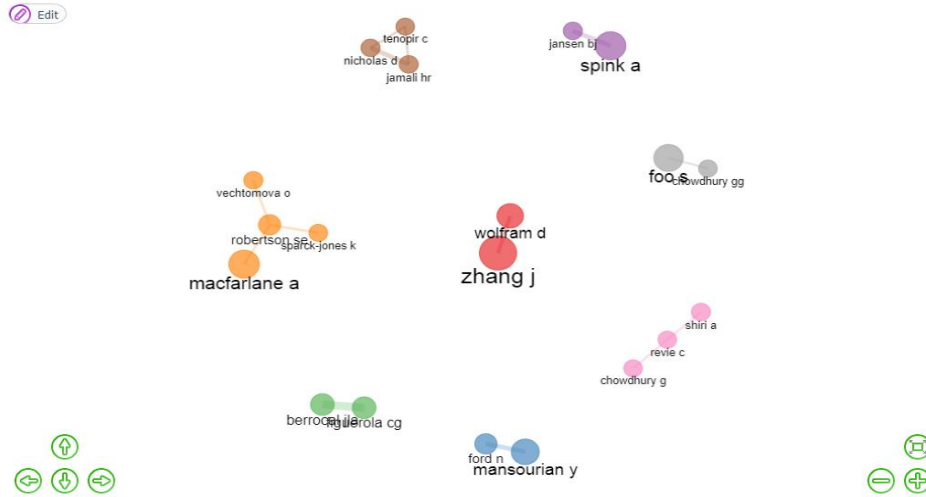
NODO	CLUSTER	BETWEENNESS	CLOSENESS	PAGERANK
Saracevic, T. 1975	3	138,09	0,014	0,019
Jansen, BJ. 2000	1	123,38	0,012	0,021
Saracevic, T. 1988	1	92,78	0,012	0,020
Belkin ,NJ. 1982	1	88,12	0,012	0,021
Bates, MJ. 1989	1	87,88	0,011	0,021
Spink, A. 1998	3	72,92	0,013	0,019
Spink, A. 2001	1	66,14	0,012	0,021
Ingwersen, A. 2005	1	56,35	0,012	0,020
Harter, SP. 1992	3	55,56	0,012	0,019
Ingwersen, P. 1996	1	50,73	0,012	0,021

Figura 31. Red de Cocitación

Esta estructura muestra la relación que existe entre unos trabajos y otros en base a la citación de esos trabajos.

### 3.8 ESTRUCTURA SOCIAL

Este tipo de estructura muestra cómo se relacionan los autores de esos trabajos, las instituciones a las que pertenecen esos autores y los países de los que proceden. En definitiva como interactúan unos con otros.



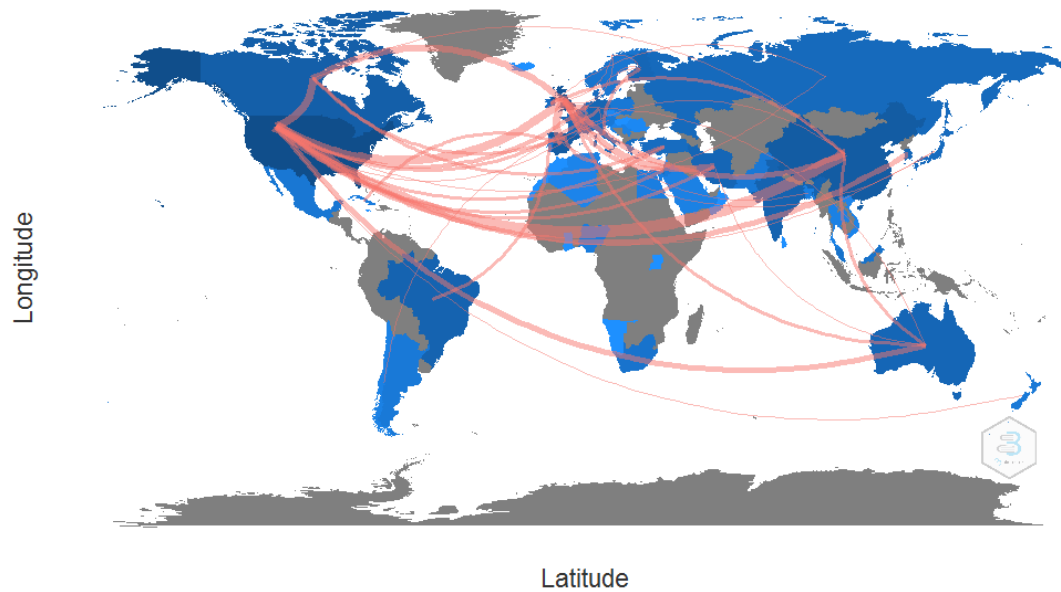
NODO	CLUSTER	BETWEENNESS	CLOSENESS	PAGERANK
Zhang, J	1	0	0,003	0,05
Wólffram, D	1	0	0,003	0,05
Mansourian, Y	2	0	0,003	0,05
Ford, N	2	0	0,003	0,05
Berrocal, Jla	3	0	0,003	0,05
Figuerola, Cg	3	0	0,003	0,05
Spink, A	4	0	0,003	0,05
Jansen, Bj	4	0	0,003	0,05
Macfarlane, A	5	0	0,003	0,03
Vechtomova, O	5	0	0,003	0,03

Figura 32. Red de Colaboración entre autores

En la figura 32 observamos como existe una retroalimentación entre los autores y sus trabajos.

### 3.8.1 Mapa de colaboración

## Country Collaboration Map



54

DESDE	HACIA	FRECIENCIA
USA	CHINA	26
USA	UNITED KINGDOM	14
UNITED KINGDOM	SPAIN	12
USA	CANADA	12
UNITED KINGDOM	CANADA	10
UNITED KINGDOM	IRAN	8
UNITED KINGDOM	CHINA	7
USA	KOREA	7
UNITED KINGDOM	NETHERLANDS	6
UNITED KINGDOM	SWITZERLAND	6

Figura 33. Mapa de Países Colaboradores

En este gráfico observamos como interactúan los autores entre ellos colaborando aunque publiquen en distintos países.

## 4. CONCLUSIONES

Sin duda, la Información de Recuperación es un área que está en expansión. Esto se debe a que los usuarios cada vez necesitan información más precisa de las materias que investigan, y debido a la cantidad de información que se produce cada minuto, no toda la que se les ofrece es buena.

Los autores, conscientes de esta necesidad, han hecho una gran labor de investigación que queda plasmada en sus publicaciones. Y esto no ha hecho más que empezar y promete dilatarse en el tiempo.

Cabe destacar la labor de cooperación entre autores, instituciones y países. Esto se demuestra cuando se observan los índices de cocitación que tienen entre ellos.

Pero toda esta labor investigadora y de publicación se hará aún más grande si los centros de información y documentación la valoran y la llevan al alza.

Como reflexión final, todas los profesionales que de alguna manera se dedican, o se dedicarán en el futuro, a la difusión de la información y documentación quizás deberían plantearse como extraer todo el jugo al campo de la Recuperación de la Información y como hacerse especialistas en ofrecer a los usuarios documentos relevantes, de calidad y adecuados a las necesidades que expresan.

Y sin duda, cualquier centro de información y documentación que quiera estar a la vanguardia de la RI, se debería plantear si es necesario adquirir un software como Biblioshiny para poder ofrecer a sus usuarios esa información de calidad.

## BIBLIOGRAFIA

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, Retrieved from <https://www.bibliometrix.org/>

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* ACM press New York.

Guzman, M., & Verstappen, B. (2002). *¿Qué es la documentación?* HURIDOCS.

Rijsbergen, C. J. (1999). *Information retrieval* Retrieved from <http://www.dcs.gla.ac.uk/~iain/keith/>

Vilares, J. Introducción a la recuperación de la información. Retrieved from <http://biblioteca.udgvirtual.udg.mx/jspui/handle/123456789/3677>