

Introducción a la documentación científica

Tomás Baiget, Jefe de Proyectos de Sistemas de Información
Instituto de Estadística de Cataluña (Idescat).

Salamanca, 3 de septiembre de 2007

Definiciones

Documento científico:

- Resultado del trabajo intelectual de investigadores y estudiosos que comunican sus reflexiones o descubrimientos registrándolos físicamente, permitiendo así la evaluación de sus pares.

Documentación científica:

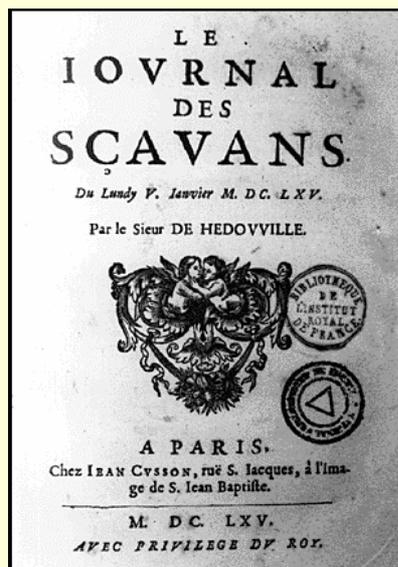
- Actividad que se ocupa de recoger y tratar los conocimientos, almacenar la información en documentos y facilitar su recuperación para contribuir a la difusión de los mismos.

- Para que algo sea **ciencia** tiene que estar escrito, pues así otras personas puedan repetir la investigación.
- El método científico obliga a hacer un rastreo de todo lo escrito antes de empezar una nueva investigación.
- Lo que no está documentado acaba desapareciendo. No es ciencia.

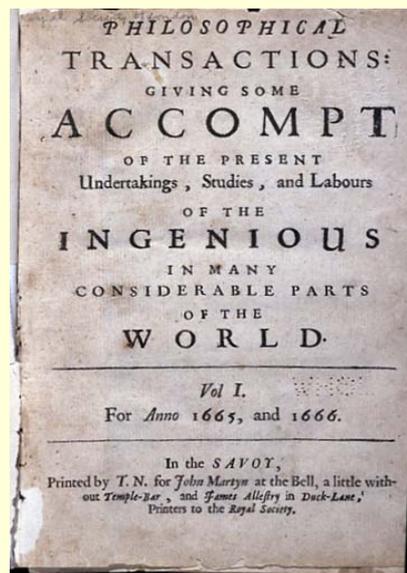
5 de enero

1665

6 de marzo

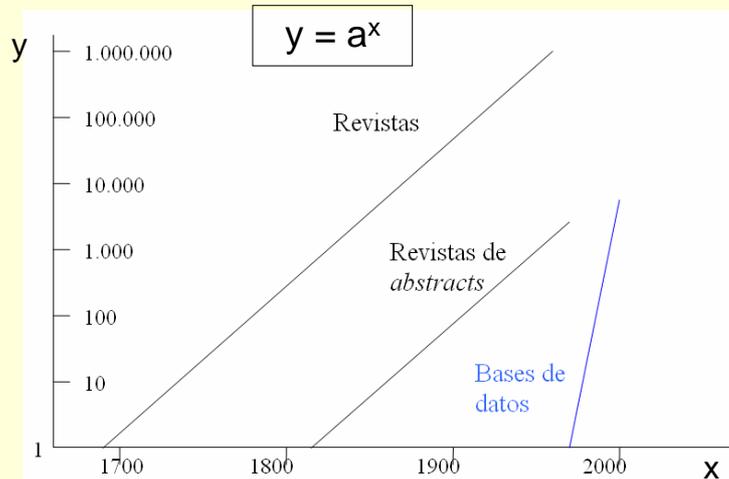


Denis de Sallo



Robert Oldenburg

Explosión de la información



Ejemplo de *abstract*

EMBASE No: 2007399779

Cochlear implants in obliterated or ossified cochleas
Implantes cocleares en cocleas obliteradas o semiosificadas
 Del Cañizo Alvarez A,

Servicio de ORL, Hospital Clinico Universitario de Salamanca, Po de San Vicente 58-182, 37002 Salamanca, **Author Email:** santaorl@usal.es

Acta Otorrinolaringologica Española (ACTA OTORRINOLARINGOL. ESP.)

(Spain) 01 AUG 2005 , 56/7 (322-327)

CODEN: AOTEA **ISSN:** 0001-6519

Document Type: Journal ; Article

Language: SPANISH; ENGLISH **Summary Language:** ENGLISH; SPANISH

Number Of References: 33

Our last cases of cochlear implants in obliterated or ossified cochleas are presented, and we had described the surgical technique for double-array or compressed system. In our opinion the best surgical method is the combined approach (endaural with aticotomy view and posterior tympanotomy) for increased promontory view. This technique provides the advantage that it always preserve the posterior wall of the external ear canal, and the array is positioned easily.

MEDICAL DESCRIPTORS:

* inner ear disease; adult; aged; article; case report; cochlea prosthesis; cochlea; electrode; female; human; male; surgical technique

SECTION HEADINGS:

011 Otorhinolaryngology
 027 Biophysics, Bioengineering and Medical Instrumentation

Metadatos

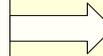
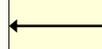
Sistema de recuperación de información (SRI)

Para obtener buenos resultados, los usuarios del SRI deben usar la misma **clasificación** y/o el mismo **vocabulario** de indización que usaron los que introdujeron los documentos

Indización de documentos



Indización de preguntas



Documentos respuesta

thesaurus



MEDICAL DESCRIPTORS:

* inner ear disease; adult; aged; article; case report; cochlea prosthesis; cochlea; electrode; female; human; male; surgical technique

SECTION HEADINGS:

011 Otorhinolaryngology
027 Biophysics, Bioengineering and Medical Instrumentation



esquema de clasificación

Clasificaciones sistemáticas o esquemas de clasificación

En la época del papel, tanto las revistas de resúmenes como las bibliotecas necesitaban sistemas de clasificación con **encabezamientos de materia** generales.

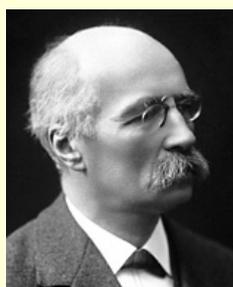
En la época informatizada esos encabezamientos se conservan porque siguen siendo útiles.

Clasificaciones bibliotecarias



Melvil Dewey
(1851-1931)

**Dewey Decimal
Classification (DDC)**



Henri la Fontaine
(1854-1943)

**Clasificación Decimal Universal
(CDU)**



Paul Otlet
(1868-1944)

Sólo se usan en bibliotecas

Ejemplo de CDU

6 Ciencias Aplicadas. Medicina. Técnica.
61 Medicina.
611 Anatomía.
612 Fisiología.
613 Higiene. Sanidad privada.
613.2 Higiene de la alimentación. Dietética.
613.4 Cuidado del cuerpo.
613.5 Higiene de la vivienda.
613.6 Higiene industrial y profesional.
613.7 Higiene del descanso.
613.81 Bebidas alcohólicas

Esquemas de clasificación de las revistas de resúmenes y bases de datos

- Al ser más especializadas, cada revista de resúmenes se hizo su propio sistema de clasificación adaptado a su temática
- Y sobre todo hicieron **vocabularios controlados** (con **descriptores**) y **thesaurus** (tesauros) especializados

Ejemplo de thesaurus: **Agrovoc, FAO**

Air pollution

BT: Pollution
RT: Atmosphere
RT: Greenhouse effect

Pollution

NT: Acid deposition
NT: Air pollution
NT: Nonpoint pollution
NT: Sediment pollution
NT: Water pollution
RT: Environmental degradation
RT: Pollutants
RT: Pesticides

BT = broader term

RT = related term

NT = narrower term

UF = used for

USE = use

Caso de integración de thesaurus y clasificación:

Medical Subject Headings (MeSH)

Anatomy	A
Body Regions	A01
Abdomen	A01.047
Abdominal Cavity	A01.047.025
Peritoneum	A01.047.025.600
Douglas' Pouch	A01.047.025.600.225
Mesentery	A01.047.025.600.451
Mesocolon	A01.047.025.600.451.535
Omentum	A01.047.025.600.573
Peritoneal Cavity	A01.047.025.600.678

Grandes bases de datos científicas

- Engineering Index (1875-) = [Compendex](#)
- Index Medicus (1879-) = [Medline](#)
- Physics (etc.) Abstracts (1898-) = [Inspec](#)
- Chemical Abstracts (1907-) = [CA Search](#)
- Biological Abstracts (1926-) = [Biosis](#)
- Excerpta Medica (1946-) = [EMBase](#)
- Sci.Citation Index (1958-) = [SCI Thomson-ISI](#)
- Derwent (1963-) = [World Patent Index](#)
- Scopus, Elsevier (1998-)

Resumen hasta aquí:

- Sin documentación no hay ciencia
- “Explosión de información”
[revistas](#) → [revistas de abstracts](#) → [bdd](#)s
- Catalogación e indización:
 - Clasificaciones sistemáticas
Generales ([CDU...](#)) y especializadas
 - Tesauros especializados de cada bdd

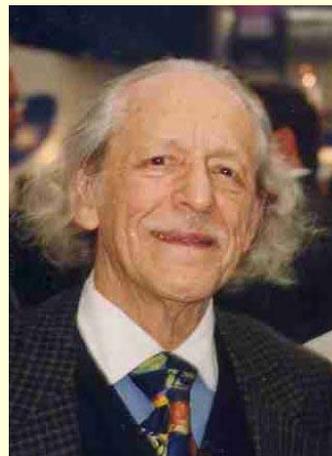
Ahora vamos a hablar de:

- *Thomson ISI (Science Citation Index)*
- Importancia del idioma inglés
- *IraLIS* (registro de autores)
- Movimiento *open access*

...e intentaremos no hablar mucho de 

Eugene Garfield, o cómo hacerse millonario de una forma “científica”

- En 1958 fundó el *Institute for Scientific Information (ISI)*, Philadelphia.
- Gran éxito comercial de los *Current Contents* (sumarios actuales)



- **Garfield** quería hacer bases de datos de referencias bibliográficas baratas, sin indización y sin resumen.
- Para “compensar” esto, añadió a las referencias las citas que los autores ponían a pie de página y al final.
- Para rentabilizar la producción, escogió un grupo reducido de las mejores revistas. Pero ¿en qué basarse?

- **Garfield** pensó que las revistas mejores eran las que los autores citaban más.
- A partir de aquí fue creando sus teorías:
 1. **Indización por citas**
 2. **Frentes de investigación**
 3. **Factor de impacto**

1. Indización por citas

- Si tenemos un artículo interesante que cita a determinados autores, otros artículos que citen a los mismos autores seguramente también serán de nuestro interés.

P. ej., buscamos artículos sobre la indización automática, y sabemos que **Gerard Salton** fue uno de sus padres:

Hacemos una búsqueda por autor citado (CA=) en el Science Citation Index:

```
Welcome to DialogClassic Web(tm)
? B34; e ca=salton g
```

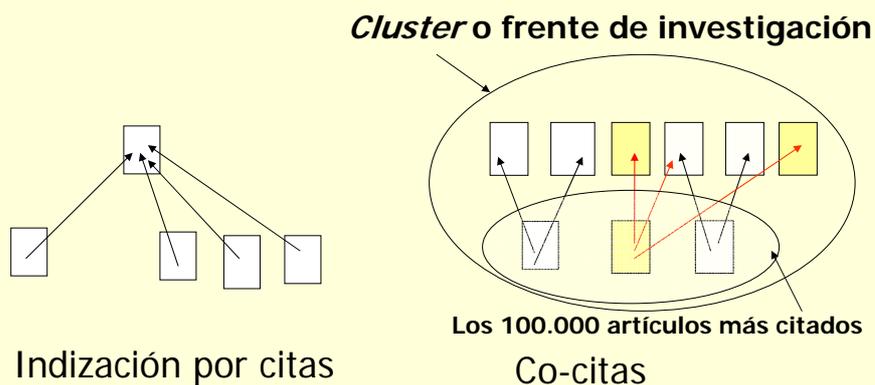
```
Ref      Items  Index-term
E3       2897  *CA=SALTON G

S E3     S1      2897  CA='SALTON G'
```

```
? T1/6/1-6
```

- 1 Semantic matching for mathematical services
- 2 Higher order feature selection for text classification
- 3 A study of interface support mechanisms for interactive information retrieval
- 4 Automatic thesaurus development: Term extraction from title metadata
- 5 Mixing semantic networks and conceptual vectors application to hyperonymy
- 6 Ontology construction for information classification

2. Frentes de investigación



Los 10.000 "frentes de investigación" definidos por el ISI, de hecho son descriptores para buscar en su base de datos.

3. Factor de impacto 2006 de una revista:

$$FI = \frac{\text{Citas hechas durante 2006 a los artículos de 2005 y 2004}}{\text{Número de artículos publicados en los años 2005 y 2004}}$$

ISI Web of KnowledgeSM Journal Citation Reports[®] HOME LOG OUT

Journal Citation Reports[®]

WELCOME ? HELP 2004 JCR Social Science Edition

Journal Summary List [Journal Title Changes](#)

Journals from: **subject categories** INFORMATION SCIENCE & LIBRARY SCIENCE [VIEW CATEGORY SUMMARY LIST](#)

Sorted by: SORT AGAIN

Journals 1 - 20 (of 54) Page 1 of 3

MARK ALL UPDATE MARKED LIST *Ranking is based on your journal and sort selections.*

Mark	Rank	Abbreviated Journal Title <i>(linked to journal information)</i>	ISSN	Total Cites	Impact Factor	Immediacy Index	Articles	Cited Half-life
<input type="checkbox"/>	1	ANNU REV INFORM SCI	0066-4200	296	4.292	0.333	12	5.7
<input type="checkbox"/>	2	INFORM SYST RES	1047-7047	615	3.512	0.100	20	5.8
<input type="checkbox"/>	3	MIS QUART	0276-7783	1869	2.884	0.333	24	9.7
<input type="checkbox"/>	4	J AM MED INFORM ASSN	1067-5027	1468	2.612	0.918	61	4.7
<input type="checkbox"/>	5	J AM SOC INF SCI TEC	1532-2882	2254	2.086	0.216	97	6.2
<input type="checkbox"/>	6	INFORM MANAGE-AMSTER	0378-7206	1013	1.815	0.056	72	5.4
<input type="checkbox"/>	7	J DOC	0022-0418	654	1.542	0.269	26	7.2
<input type="checkbox"/>	8	INFORM PROCESS MANAG	0306-4573	992	1.295	0.185	54	7.8
<input type="checkbox"/>	9	J MANAGE INFORM SYST	0742-1222	944	1.271	0.171	35	6.9
<input type="checkbox"/>	10	INT J GEOGR INF SCI	1365-8816	899	1.234	0.135	37	7.4
<input type="checkbox"/>	11	COLL RES LIBR	0010-0870	384	1.159	0.038	26	5.8

Thomson ISI – Web of Science

WoS está compuesto de 3 bases de datos principales:

Science Citation Index (1900-)

Social Science Citation Index (1956-)

Arts & Humanities Index (1975-)

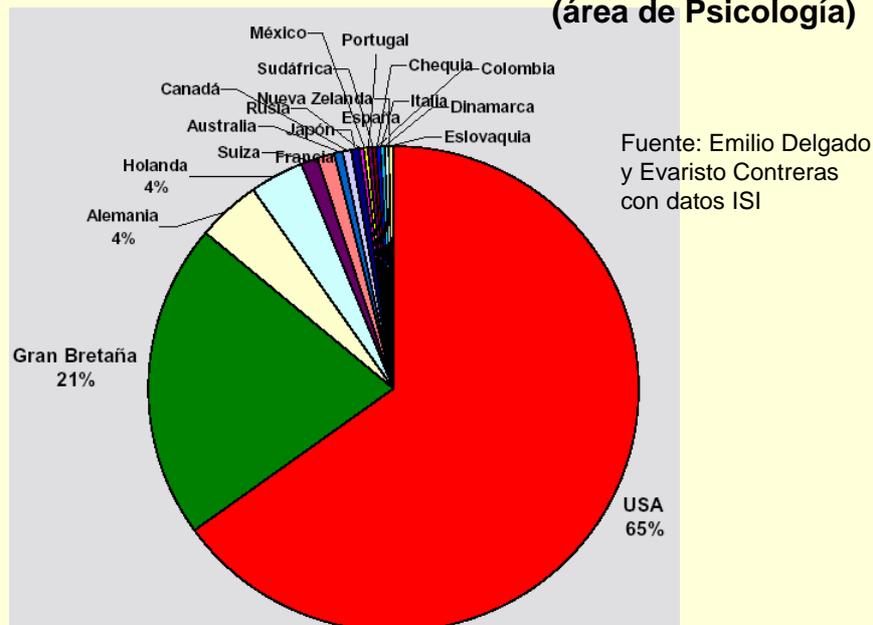
que suman unos 9.000 títulos de las mejores revistas del mundo, pero:

-sesgo anglosajón

-desequilibrio entre las ciencias

Sesgo geográfico del WoS

(área de Psicología)



ISI WoS (Web of Science)

	Revistas incluidas	Nuevos registros semanales	Nuevas referencias citadas semanalmente	Disciplinas incluidas	Revistas por disciplina
SCI	6.126	22.200	420.600	27	6126 / 27 = 226,8
SSCI	1.802	3.000	70.600	32	1802 / 32 = 56,3
AHI	1.136	1.800	15.500	20	1136 / 20 = 56,8

Cada investigador STM tiene 226 oportunidades de publicar un artículo en una revista ISI, frente a las 56 oportunidades en ciencias sociales e humanidades

Alternativas a Thomson ISI

SCOPUS Thomas target is logged in [Logout](#)

Search Sources My Alerts My List My Profile Help Scopus Labs

NEW: Find out about the latest enhancements in Scopus. [Click here.](#) Brought to you by Scopus Team Library catalogue

Basic Search **Author Search** Advanced Search Search Tips

Search for: in

E.g., "heart attack" AND stress

AND in

Limit to: **Date Range** (inclusive)
 Published to
 Added to Scopus in the last days

Document Type

Subject Areas ⓘ
 Life Sciences (> 3,400 titles)
 Health Sciences (> 5,300 titles)
Includes 100% Medline coverage
 Physical Sciences (> 5,500 titles)
 Social Sciences (> 2,800 titles)

Antropología	Educación
Biblioteconomía y Documentación	Geografía
Ciencia política y de la administración	Psicología
Comunicación	Sociología
Economía	Urbanismo

Financiado

Qué es
Por qué es necesario
Como se ha confeccionado
A quién puede interesar
Quién lo elabora
Más información
 English



Dirección General de Universidades
Proyectos EA2003-086, EA2004-0119,
EA2005-0068, EA2006-0018, EA2006-
0027
Dirección General de Investigación
Proyecto SEJ2004-08027/SOCI

Estadísticas Generales

Estadísticas de accesos: 143845



EC³

Grupo de investigación
Evaluación de la ciencia y de la comunicación científica
Universidad de Granada



El idioma



- El idioma determina el área geográfica donde el fruto de nuestra investigación puede tener valor.
- El inglés ya es la lengua de la comunicación científica mundial → Siempre que podamos tenemos que hacer traducir nuestro artículo.
- Todo lo que no sea inglés sólo se puede considerar divulgación, nada nuevo, ya que casi no afecta al **estado de la técnica**.

La influencia del inglés nos obliga a cambiar la forma **cómo firmamos los hispanos**. Si un autor firma, p. ej., **Fernanda Peset Mancebo**, su trabajo aparecerá en el ISI como **Mancebo, FP**

<http://iralis.org>

The screenshot shows the Iralis website interface. At the top left is the Iralis logo, followed by the text 'International Registry for Authors in Library and Information Science'. Below this is a navigation menu with items like 'Why Iralis', 'How it works', 'Success assessment', etc. The main content area features an article titled 'Why Iralis' with a sub-header 'Submitted by editor on Sun, 2007-01-21 17:10.'. The article text explains the purpose of Iralis and discusses naming conventions in Spanish and other languages.

La espiral de revistas

Causas:

1. Redes y consorcios de bibliotecas
→ **compartir suscripciones**
2. Facilidad para fotocopiar, y para obtener copias-e de los artículos originales
→ **pasar del *just-in-case* al *just-in-time***

Ante las cancelaciones, las editoriales incrementan los precios, lo cual provoca aún más cancelaciones.

Movimiento *acceso abierto*

El encarecimiento de las revistas hizo surgir el movimiento *open access* (OA), que quiere mentalizar a los autores para que:

- depositen sus trabajos en webs (“**vía verde**”)
- publiquen en revistas gratuitas, costeadas por los propios autores o por sus instituciones (“**vía oro**” al OA)

Editores OA:

- BioMed Central
- Public Library of Science



Stevan Harnad,
Univ. of Southampton

Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)

Normas sobre cómo deben hacerse los repositorios y cómo deben indizarse los documentos (usando **Dublin Core**), para que puedan hacerse búsquedas.

Dos objetivos generales:

- Conservar los documentos
- Difundirlos en OA

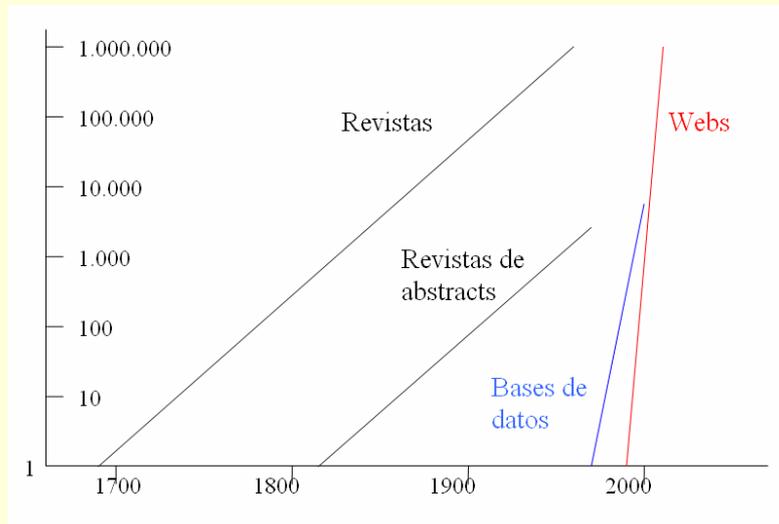
Algunos repositorios

- arXiv física, matemáticas 350.000
- PubMed Central medicina
- RePec economía
- E-LIS bibls., documentación 6.500
- NASA ADS astrofísica
- Ideas economía 250.000
- SciELO bio-medicina

Nuevas facetas de la documentación científica

- publicación electrónica propia
- web 2.0 informática social, blogs, wikis
- gestión del conocimiento
- indicadores de ciencia y tecnología (webmetría)
- *data mining* y *knowledge discovery*

Con internet continúa la explosión de información...



Y ahora estamos yendo a remolque de lo que desarrollan por la Red...

G-Book, G-DeskTop, G-Maps, G-Earth,
G-Scholar, G-Analytics, G-Translation, G-Blogger
G-Spelling, G-Images & videos, G-Calendar...

¿Indización manual vs automatizada? 😊

¿Es necesaria tanta catalogación en las bibliotecas?

Bibliografía

Garfield, Eugene. "Research fronts". In: *Current Contents*, Oct. 10, 1994,
<http://scientific.thomson.com/free/essays/citationanalysis/fronts/>

Garfield, Eugene. "The ISI impact factor". In: *Current Contents*, Jun. 20, 1994,
<http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/>

Ortiz-Trujillano, Amalia. Introducción a la Documentación Científica. Estrategias de Búsqueda Bibliográfica.
http://same.ens.isciii.es/Fuentes%20de%20informaci%C3%B3n_CT_salud/Documentacion/tema%201/Introducci%C3%B3n_Doc_archivos/frame.htm

Peris-Bonet, R. Proyecto Docente. Valencia, 2001.
<http://www.uv.es/Ciencia/Capx.pdf> (x = 1...10)

Peset-Mancebo, Fernanda; Ortín-Pérez, Alfredo. Tendencias internacionales en el acceso libre a la documentación científica digital: OAI-Open Archives Initiative, 2005
<http://eprints.rclis.org/archive/00004131/>

Urra, Pedro. Biblioteca Virtual en Ciencias de la Salud. Habilidades informacionales: recursos, metodología y técnicas de trabajo científico. Introducción a la documentación científica. CNICM/Infomed, pedro.urr@infomed.sld.cu
<http://www.usal.es/~sabus/site%20med/descargas/Urra.pdf>

¡Muchas gracias por su atención!

¿Preguntas?



Transparencias en:

<http://eprints.rclis.org/>