



Introduction à la bibliométrie

Stéfan DARMONI

Professeur d’Informatique Médicale, Faculté de Médecine de Rouen

Département de l’Informatique et de l’Information Médicale, CHU de

Rouen

Tous droits réservés



LIMICS U1142 INSERM

Critères de qualité

Citez vos sources !!!

- INITIATION A LA
BIBLIOMETRIE

- <http://www.pasteur.fr/infosci/biblio/formation/metrie/>
- Baudoin L, Haeffner-Cavaillon N, Pinhas N, Mouchet S, Kordon C. [Bibliometric indicators realities, myth and prospective]. Med Sci (Paris). 2004 Oct;20(10):909-15.

Définition

- La biométrie se définit comme l'exploitation statistique des publications.
Cette analyse permet de rendre compte de l'activité des producteurs (chercheur, laboratoire, institut...) ou des diffuseurs (périodique, éditeur...) de l'information scientifique, tant d'un point de vue quantitatif que qualitatif.
- Bibliometrics
The use of statistical methods in the analysis of a body of literature to reveal the historical development of subject fields and patterns of authorship, publication, and use. Formerly called statistical bibliography. (from The ALA Glossary of Library and Information Science, 1983)
Year introduced in the Medline database: 1990
Number of papers indexed with this MeSH term = 9767 (Jan 2019)



Quelques applications de la bibliométrie

- évaluer le travail d'un chercheur, d'un labo (y compris auto-évaluation...) ou le définir par analyse sémantique (d'un chercheur ou d'un groupe de chercheur, d'un centre, d'un pays...)
- évaluer le fonds de périodiques d'une bibliothèque
- suivre l'évolution d'un thème de recherche
- apprécier l'impact d'un article
- évaluer la qualité d'une revue...

Moyens de la biométrie

- Les publications, pour une grande part, sont référencées dans des bases de données bibliographiques **informatisées**; ce qui permet des **recherches rapides**, selon des **critères multiples**, sur **d'importants volumes** de références.

Le choix des bases de données dont seront extraites les références est essentiel : il détermine quel type d'information va être analysé.



Moyens de la biométrie (suite)



- Les outils informatiques :
Un **tableur** peut suffire pour réaliser des analyses biométriques de base. Il faut pouvoir :
 - **importer des références** depuis une base de données
 - faire **des tris et des comptages**
 - faire des **exploitations graphiques**

L'analyse de corpus complexes (par la méthode des mots associés par exemple) nécessite des logiciels spécifiques tels que Leximappe ou Dataview.

Bases de données bibliographiques

- Contenu : articles scientifiques
- Il en existe de plusieurs sortes
- Celles dont l'objectif est de fournir les résumés des articles => Medline, Embase
- Celles dont l'objectif est la bibliométrie => ISI
- Principaux critères de jugement :
 - la **couverture** (domaine, zone géographique, type de publications, langues, nombre de sources dépouillées...)
 - la **forme** et l'**organisation** de l'information (organisation en champs, nombre de champs, type d'indexation...)
 - Chaque base de données possède ses propres spécificités. Les références qui y sont regroupées sont sélectionnées selon des critères variables, en fonction des buts visés et des publics auxquels elles sont destinées.

Bases de données bibliométriques

ISI Institute for Scientific Information

<http://www.isinet.com/>

Un Gourou : Eugene Garfield



Science Citation Index SCI, devenu Web of
Science (WoS)

Journal Citation Report JCR



SCI Science Citation Index Expanded

Science Citation Index Expanded is a multidisciplinary index to the journal literature of the sciences.

It fully indexes 5,900 major journals across 150 scientific disciplines—that's 2,100 more journals than the print and CD-ROM versions of the *SCI*.

The *Science Citation Index Expanded* includes all cited references captured from indexed articles.

- In addition, the *Science Citation Index Expanded*: Provides access to current information and retrospective data **from 1945** forward. Note that you may not have access to all data years.
- Averages 19,000 new records per week.
- Includes approximately 423,000 new cited references per week.
- Some of the disciplines covered include: Agriculture, Neuroscience, Astronomy, Oncology, Biochemistry, Pediatrics, Biology, Pharmacology, Biotechnology, Physics, Chemistry, Plant Sciences, **Computer Science**, Psychiatry, Materials Science, Surgery, Mathematics, Veterinary Science, Medicine, Zoology

JCR

- *Journal Citation Reports* is a comprehensive and unique resource tool that allows you to evaluate and compare journals using citation data drawn from over 7,000 scholarly and technical journals from more than 3,300 publishers in over 60 countries.
- It is a source of citation data on journals, and includes virtually all specialties in the areas of science, technology, and social sciences.
- *Journal Citation Reports* can show you the:
 - Most frequently cited journals in a field
 - Highest impact journals in a field
 - Hottest journals in a field
 - Leading journals in a field
 - Most published articles in a field

Impact Factor (IF) Facteur d'impact (FI)

- **Le facteur d'impact** d'un journal, indicateur le plus utilisé, mesure à un temps t le nombre de citations relevé au cours des années $t-1$ et $t-2$ pour les articles publiés durant ces mêmes années.
- Ainsi, le facteur d'impact de *Nature* en 2002 représente le quotient du nombre de citations en 2002 et du nombre d'articles publiés en 2000 et 2001, soit $(33448 + 25955)/(1013 + 939) = 30,432$. Il équivaut donc à la moyenne annuelle rétrospective, calculée sur deux ans et donc à relativement court terme, de la «consommation» annuelle moyenne des publications d'une revue. À l'origine, cet indicateur était utilisé pour la sélection des journaux dans les *Current Contents**.

Critiques de l'IF

- The impact factor (IF) published by the Institute for Scientific Information (ISI):
 - quantifies the influence of a periodical on secondary publications and
 - is very commonly used not only to rank and evaluate journals but also for academic promotion or for the selection of research grant applications.
- IF is marred by numerous limitations:
 - scale of IF varies widely between scientific fields and medical specialities
 - a given IF is not, per se, a good indicator of scientific value.
 - there is no correlation between the citation frequency of a given paper and the impact factor of the journal in which it has been published

Indice d'immédiateté

- mesure la rapidité avec laquelle les articles d'un journal donné sont cités.
- Il représente le quotient du nombre de citations sur le nombre d'articles publiés la même année. Par exemple, l'indice d'immédiateté de *Nature* en 2002 représente le quotient du nombre de citations en 2002 et du nombre d'articles publiés en 2002 (soit $6671/889=7,504$).
- Pour la grande majorité des revues scientifiques, la valeur de cet indice se situe entre 0 et 1. Pour seulement deux revues (*Nature* et *Journal of Experimental Medicine*), il est supérieur à 7. Cet indicateur permet de repérer les journaux qui concentrent les publications «chaudes».

Demi-vie de citation

- La demi-vie de citation rend compte de la longévité des articles d'un journal, et donc de la persistance de sa notoriété à long terme.
- Il se définit par la durée au-delà de laquelle la moitié des articles de l'année de référence cesse d'être citée.
- À titre d'exemple, cet indice est de 6,9 (ans) pour *Nature* en 2002.

**ISI Web of KnowledgeSM**

— Products & Features —

GO

HOME

LOG OUT

Welcome to *ISI Web of Knowledge*... transforming research

ISI Web of Knowledge is a dynamic, fully integrated environment that provides one source for high quality content and tools to access, analyze, and manage research information. Comments or ideas? [Please click here to send feedback to the development staff.](#)

More information
for new users

[click here](#)
[More information](#) | [Notices](#) | [Help](#) | [Tutorial](#)
CrossSearch Example: quark* and spin

Enter a topic

SEARCH

[More search options and
External Collections](#)

Searchable Database Products

Web of Science

Science Citation Index Expanded
Index Chemicus
Current Chemical Reactions
Social Sciences Citation Index
Arts & Humanities Citation Index

GO

Current Contents Connect

Current journals, Web sites, and
books – updated daily

GO

ISI Proceedings

International conferences and
meetings

GO

Derwent Innovations Index

International patents

GO

BIOSIS Previews

Biology and Life Sciences

GO

Analytical Tools

Journal Citation Reports

GO

Journal performance metrics,
including Impact Factor

Essential Science Indicators

GO

Scientific performance measures

Other Resources

ISI HighlyCited.com

GO

Author biographies and bibliographies

www.thomsonisi.com

GO

Thomson ISI's Web site

Please [register](#) for
more features i

Sign In:

E-mail Address:

Password:

[Forgot your password](#)

SIGN IN

Citation Alerts
[View My Cited Articles List](#)
My Journal List
[Create My Journal List and
Table of Contents Alerts](#)

Fichier Edition Affichage Favoris Outils ?

Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame>

OK Liens

ISI Web of KnowledgeSM

Web of Science

GO

HOME

LOG OUT



Science Citation Index Expanded™ (SCI EXPANDED™)

Current Chemical Reactions® (CCR®)

Index Chemicus® (IC®)

Social Sciences Citation Index® (SSCI®)

Arts & Humanities Citation Index® (A&HCI®)

[Information for New Users](#)

Select a search option:

Quick search: GO

Example: chess AND comput*

GENERAL SEARCH

Search by topic, author, journal, or address.

CITED REFERENCE SEARCH

Search for articles that cite an author or work.

STRUCTURE SEARCH

Search for reactions and compounds using chemical structures.

ADVANCED SEARCH

Search using complex queries including field tags and set combinations.

OPEN SAVED SEARCH

Open a previously saved search history.

Select database(s) and timespan:

Citation Databases:

- Science Citation Index Expanded (SCI-EXPANDED)--1945-present
 Social Sciences Citation Index (SSCI)--1956-present
 Arts & Humanities Citation Index (A&HCI)--1975-present

Chemistry Databases:

- Current Chemical Reactions (CCR-EXPANDED)--1986-present
(includes Institut National de la Propriete Industrielle structure data back to 1840)
 Index Chemicus (IC)--1993-present

Latest (updated November 13, 2004)
 Year
 From to (default is all years)

To remember these settings, first [sign in or register](#).

[Fichier](#) [Edition](#) [Affichage](#) [Favoris](#) [Outils](#) ?Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame>

OK Liens

ISI Web of KnowledgeSM

Web of Science

GO

HOME

LOG OUT

Web of Science®



Cited Reference Search

[>> View your search history/combine sets](#)

Selected database(s) and timespan:

Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=1945-2004

CHANGE SETTINGS

Find the citations to a person's work by entering the person's name, the work's source, and/or publication year.

[View our Cited Reference Searching tutorial](#)[SEARCH](#) [CLEAR](#)**CITED AUTHOR:** Enter the name of the cited author (see [cited author index](#)).

Example: O'BRIAN C* OR OBRIAN C*

CITED WORK: Enter the abbreviated journal/book title in which the work appeared, a patent number, or another work (see [cited work index](#) or [view the Thomson ISI list of journal abbreviations](#)).

Example: J Comput Appl Math*

CITED YEAR(S): Enter year, or range of years, the cited work was published.

Examples: 1943 or 1943-1945

[SEARCH](#) [CLEAR](#)

ISI Web of KnowledgeSM

Web of Science

GO

HOME

LOG OUT

Select only those cited references you want to include,
then click FINISH SEARCH.

(Hint: Look for variants. Papers are sometimes cited incorrectly.)

[FINISH SEARCH >>](#)

View the articles that cite the selected references.

The completed search will be added to the search history.

[\(Limit by language and document type\)](#)

CITED REFERENCE INDEX

References 1 -- 20

Go to Page: of 4 GO

◀◀◀ [1 | 2 | 3 | 4] ▶▶▶

[SELECT PAGE](#)[SELECT ALL*](#)[CLEAR ALL](#)

or select specific references from the list.

When desired references have been selected from all pages, click FINISH SEARCH to complete your search.

Select	Times Cited**	Cited Author	Cited Work	Year	Volume	Page	Article ID	View Record
<input type="checkbox"/>	2	ZWEIGENBAUM P	13TH P COLING LUGANO	1990		82		
<input type="checkbox"/>	1	ZWEIGENBAUM P	17 AIMMENELAS	1995				
<input type="checkbox"/>	1	ZWEIGENBAUM P	17 MENELAS	1995				
<input type="checkbox"/>	1	ZWEIGENBAUM P	19 S COMP APPL MED S	1995		357		
<input type="checkbox"/>	1	ZWEIGENBAUM P	ACL WORKSH NAT LANG	2002				
<input type="checkbox"/>	1	ZWEIGENBAUM P	ACT CINQ JOURN FRANC	1994		23		
<input type="checkbox"/>	1	ZWEIGENBAUM P	ACTES TALN 2003	2003		285		
<input type="checkbox"/>	1	ZWEIGENBAUM P	AIM 92	1992		215		
<input type="checkbox"/>	1	ZWEIGENBAUM P	AIM9 MEN DEL REP	1993				
<input type="checkbox"/>	3	ZWEIGENBAUM P	AMIA 97	1997		590		
<input type="checkbox"/>	3	...Zweigenbaum P	APPL ARTIF INTELL	1994	8	425		View record
<input type="checkbox"/>	1	ZWEIGENBAUM P	ARTIF INTELL	1999		416		
<input type="checkbox"/>	2	ZWEIGENBAUM P	COLING 90	1990		82		
<input type="checkbox"/>	1	ZWEIGENBAUM P	COMPREHENSION AUTOMA	1987				
<input type="checkbox"/>	28	ZWEIGENBAUM P	COMPUT METH PROG BIO	1994	45	117		View record
<input type="checkbox"/>	1	ZWEIGENBAUM P	ELRA NEWSLETTER	2001	6	8		

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Média

Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame> OK Liens

ISI Web of KnowledgeSM Web of Science GO HOME LOG OUT

Web of Science®

WELCOME HELP GENERAL SEARCH CITED REF SEARCH STRUCTURE SEARCH SEARCH HISTORY ADVANCED SEARCH

Full Record

[<< Return to Cited Reference Index](#)

Record 1 of 1

Title: SEMANTIC ANALYSIS AND IN-DEPTH UNDERSTANDING OF TECHNICAL TEXTS

Author(s): CAVAZZA M, ZWEIGENBAUM P

Source: APPLIED ARTIFICIAL INTELLIGENCE 8 (3): 425-453 JUL-SEP 1994

Document Type: Article

Language: English

Cited References: 37 **Times Cited:** 3 [FIND RELATED RECORDS](#)

Abstract: We explored the problem of achieving in-depth understanding of natural language sentences from narrative technical reports through knowledge-based free text understanding. We rely on the assumption that texts in an expert domain convey much implicit information, which can be recovered by building and reasoning on a model of the situation described with the help of both linguistic and detailed world knowledge. We describe a two-step approach to semantic analysis: the first step assembles a conceptual representation of a sentence and deals with linguistic issues; the second step actually builds and runs the situational model and is totally dedicated to representation and inference. We evaluated this approach by designing a research prototype that processes sentences from clinical narratives in a medical specialty. This proto-type was fully implemented and was tested on actual sentences. We hereby give a detailed account of this implementation as well as the first results obtained.

KeyWords Plus: MENTAL MODELS

Addresses: INSERM, U194, DEPT INTELLIGENCE ARTIFICIELLE & MED, PARIS, F-75005 FRANCE
AP HP, SERV INFORMAT MED, PARIS, FRANCE

Publisher: TAYLOR & FRANCIS, 1900 FROST ROAD, SUITE 101, BRISTOL, PA 19007-1598

Subject Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE; ENGINEERING, ELECTRICAL & ELECTRONIC

IDS Number: PD602

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Média Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame> OK Liens

ISI Web of KnowledgeSM Web of Science GO HOME LOG OUT

Web of Science®

WELCOME HELP GENERAL SEARCH CITED REF SEARCH STRUCTURE SEARCH SEARCH HISTORY ADVANCED SEARCH SEARCH RESULTS

Cited References

SEMANTIC ANALYSIS AND IN-DEPTH UNDERSTANDING OF TECHNICAL TEXTS
CAVAZZA M, ZWEIGENBAUM P
APPLIED ARTIFICIAL INTELLIGENCE
8 (3): 425-453 JUL-SEP 1994

The following documents are bibliographic references cited by the above article:

[FIND RELATED RECORDS](#)

Clear the checkbox to the left of an item if you do not want to retrieve articles that cited the item when finding Related Records.

<input type="checkbox"/> CLEAR ALL	Cited Author	Cited Work	Year	Volume	Page	Article ID	View Record
<input checked="" type="checkbox"/>	ALLEN JF	NATURAL LANGUAGE UND	1987				
<input checked="" type="checkbox"/>	BOUAUD J	13TH P INT JOINT C A	1993		496		
<input checked="" type="checkbox"/>	BOWER GH	SCIENCE	1990	247	44		View record
<input checked="" type="checkbox"/>	CAVAZZA M	IN PRESS COMPUTATION	1994				
<input checked="" type="checkbox"/>	CAVAZZA M	THESIS U PARIS 7	1991				
<input checked="" type="checkbox"/>	CHAFFIN R	RELATIONAL MODELS LE	1988		289		
<input checked="" type="checkbox"/>	CHANDRASEKARAN B	ARTIF INTELL	1989	1	29		
<input checked="" type="checkbox"/>	DORE L	THESIS U PARIS 13	1992				
<input checked="" type="checkbox"/>	FILLMORE CJ	UNIVERSALS LINGUIST	1968		1		
<input checked="" type="checkbox"/>	FORGY CL	ARTIF INTELL	1982	19	17		View record
<input checked="" type="checkbox"/>	GARNHAM A	MENTAL MODELS REPRES	1986				
<input checked="" type="checkbox"/>	GENTER D	MENTAL MODELS	1983				
<input checked="" type="checkbox"/>	GRISHMAN R	ANAL LANGUAGE RESTRI	1986				
<input checked="" type="checkbox"/>	GUHA RV	APPL ARTIF INTELL	1991	5	45		View record
<input checked="" type="checkbox"/>	HIRST G	SEMANTIC INTERPRETAT	1987				

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris Média * Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame> OK Liens

ISI Web of KnowledgeSM Web of Science GO HOME LOG OUT

Web of Science®



Citing Articles--Summary

[SEMANTIC ANALYSIS AND IN-DEPTH UNDERSTANDING OF TECHNICAL TEXTS](#)

CAVANNA M, ZWEIGENBAUM P

APPLIED ARTIFICIAL INTELLIGENCE

8 (3): 425-453 JUL-SEP 1994

These documents in the database cite the above record:

3 results found

Records 1 -- 3

Go to Page: of 1 GO

[<] [<<] [<] [1] [>] [>>] [>]

Use the checkboxes to select individual records for marking, then click Submit to add them to the Marked List.

- 1. Cavazza M, Palmer I
[High-level interpretation in virtual environments](#)
APPLIED ARTIFICIAL INTELLIGENCE 14 (1): 125-144 JAN 2000
- 2. Haouche-Gingins C, Charlet J
[A knowledge-level testing method](#)
INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES 49 (1): 1-20 JUL 1998
- 3. ZWEIGENBAUM P, BACHIMONT B, BOUAUD J, et al.
[ISSUES IN THE STRUCTURING AND ACQUISITION OF AN ONTOLOGY FOR MEDICAL LANGUAGE UNDERSTANDING](#)
METHODS OF INFORMATION IN MEDICINE 34 (1-2): 15-24 MAR 1995

Mark: [0 records marked]

Selected records

All records on this page

Records to

SUBMIT

You can print, save, export, e-mail, and order records after adding them to the Marked List. (The list can hold 500 records.)

Analyze Results:

ANALYZE

View rankings and histograms of the authors, journals, etc. for this set of records. (Up to 2,000 records at a time.)

[Fichier](#) [Edition](#) [Affichage](#) [Favoris](#) [Outils](#) ?Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>

OK Liens

ISI Web of KnowledgeSM

Journal Citation Reports

GO

HOME

LOG OUT

ISI JOURNAL
CITATION
reports[®]Powered by ISI Web of KnowledgeSM[Information for New Users](#)

Select a JCR edition and year:	Select an option:
<input checked="" type="radio"/> JCR Science Edition <select>2003</select>	<input checked="" type="radio"/> View a group of journals by <select>Subject Category</select>
<input type="radio"/> JCR Social Sciences Edition <select>2003</select>	<input type="radio"/> Search for a specific journal
	<input type="radio"/> View all journals
<input type="button" value="SUBMIT"/>	

This product is best viewed in 800x600 or higher resolution

[NOTICES](#)

The Notices file was last updated Thu Oct 21 14:25:11 2004

THOMSON
[Acceptable Use Policy](#)Copyright © 2004 [The Thomson Corporation](#)

[Fichier](#) [Edition](#) [Affichage](#) [Favoris](#) [Outils](#) ?Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>

OK Liens

ISI Web of KnowledgeSM

Journal Citation Reports

GO

HOME

LOG OUT

ISI JOURNAL CITATION reports®

Powered by ISI Web of KnowledgeSM

2003 JCR Science Edition

SUBJECT CATEGORY SELECTION

Select one or [more subject categories](#) from the list to filter journals.

Read subject category scope notes:

[SCOPE NOTES](#)

- MEDICAL INFORMATICS
- MEDICAL LABORATORY TECHNOLOGY
- MEDICINE, GENERAL & INTERNAL
- MEDICINE, LEGAL
- MEDICINE, RESEARCH & EXPERIMENTAL
- METALLURGY & METALLURGICAL ENGINEERING
- METEOROLOGY & ATMOSPHERIC SCIENCES
- MICROBIOLOGY
- MICROSCOPY

Sort journals by:

Journal Title

SUBMIT

[Acceptable Use Policy](#)Copyright © 2004 [The Thomson Corporation](#)

Fichier Edition Affichage Favoris Outils ?

Adresse <http://isi9.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>

OK Liens

ISI Web of KnowledgeSM

Journal Citation Reports

GO

HOME

LOG OUT

ISI JOURNAL CITATION reports®

Powered by ISI Web of KnowledgeSM

2003 JCR Science Edition

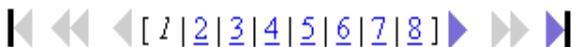
JOURNAL SUMMARY LIST

Selection: FRANCE

Sorted by: Impact Factor ▾ SORT AGAIN

JOURNAL TITLE CHANGES

Journals 1 - 20 (of 147)



Page 1 of 8

MARK ALL UPDATE MARKED LIST

Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title <i>(linked to full journal information)</i>	ISSN	2003 Total Cites	Impact Factor	Immediacy Index	2003 Articles	Cited Half-life
<input type="checkbox"/>	1	ASTRON ASTROPHYS	0004-6361	63021	3.843	1.084	1936	5.8
<input type="checkbox"/>	2	MICROBES INFECT	1286-4579	2301	3.772	0.288	160	2.9
<input type="checkbox"/>	3	BIOCHIMIE	0300-9084	3675	3.707	0.158	120	5.4
<input type="checkbox"/>	4	ESA BULL-EUR SPACE	0376-4265	1485	3.683	0.909	44	5.0
<input type="checkbox"/>	5	CHEM SENSES	0379-864X	2085	2.691	0.400	80	6.0
<input type="checkbox"/>	6	EUR PHYS JE	1292-8941	1145	2.445	0.402	169	2.4
<input type="checkbox"/>	7	GEOSTANDARD NEWSLETT	0150-5505	674	2.410	0.400	15	6.7
<input type="checkbox"/>	8	BIOL CELL	0248-4900	1302	2.159	0.091	66	9.1

PubMed/Medline

- Succès majeur des systèmes documentaires
- Base de données bibliographique la plus utilisée dans le monde
- Disponible
 - depuis 40 ans, sous forme papier
 - depuis ≈ 25 ans, sous forme électronique
 - depuis ≈ 7 ans, gratuitement sur l'Internet
- 26 millions de références en 2019
- 4.000 revues scientifiques (20% de la littérature mondiale)

Medline (suite)

- Biais US vs. Europe >> France
- Thésaurus MeSH
 - Hiérarchisation => explosion
- Principe de recherche d'information
 - « mc » sinon « all fields »
 - AIDS[MeSH] / AIDS[All fields]



Autres bases de données bibliographiques



- Embase
 - Europe
 - Thérapeutique / Toxicologie
 - Thésaurus ≠ MeSH
- Biosis
- Pascal
- BDD bibliographiques spécialisées
 - AIDSline, CancerLine, ToxLine, ...
- Recherche multi-base
 - €€€
 - Effectué par la bibliothèque médicale
 - Réservé au soin



Types d'information

- Information primaire
 - Ensemble du document en texte intégral
 - Exemple : bibliothèque virtuelle avec journaux électroniques
 - Banque d'information sur le médicament
 - Banque de RBPC
- Information secondaire
 - Partie du document +/- annoté
 - Base de données bibliographiques (Medline)
 - Catalogue de ressources (CISMeF)



Critères pour mesurer un système documentaire

	Pertinent	Non pertinent	
Documents transmis	A	B	A+B
Documents non transmis	C	D	C+D
	A+C	B+D	

Rappel = $A/A+C$; silence = $1-Rappel = C/A+C =$ faux négatifs

Précision = $A/A+B$; bruit = $1 - Précision = B/A+B =$ faux positifs



The Web Impact Factor

- WIF (Ingwersen 1998; Journal of Documentation):
 - ≡ Impact Factor (Journal Citation Reports)
 - applied to Web Sites
 - absolute WIF: nb of external pages pointing to the site
 - $aWIF(W)$ = number of external pages pointing to W
 - W : a web site; collection of sites; selection of pages in site
- nb of links:
 - indicator of the impact of the site
 - used in ranking algorithms of search engines such as Google

Use of PubMed

- Calculate the number of publications in Medline: use of the PubMed Web site
 - URL: <http://www.ncbi.nlm.nih.gov/pubmed>
- Formula:
 - Nb= Medical Informatics [MeSH Terms] and Country[Affiliation]
 - Nb(USA)=Medical informatics [MeSH Terms] and USA[Affiliation]
= 44,860

PageRank de Google

- PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value.
- In essence, Google interprets a link from page A to page B as a vote, by page A, for page B.
- But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."
- Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search.



Discussion (1/3)

- Bibliometrics = the application of quantitative research techniques to published documents
- Practical applications, for instance in evaluating journal holdings in libraries
- To apply the methods of bibliometrics to the Internet:
 - these informetric methods applied to the Web are called ‘webometrics’, or ‘cybermetrics’, or ‘infometrics’

Discussion (2/3)

- The WIF has been proposed as one possible criterion to help assess the quality of health information sites on the Internet ...
- ...but the WIF does not measure the overall quality of a Web site
- Correlation between quality of the site and quality of the medical content is still disputed
- WIFs do not measure the quality of research, but some quantity that correlates with it
- This study does not show a direct impact of a scientific society on the publications in a country

Discussion (3/3)

- Feasibility and reproducibility of the WIF
- WIF can be used over time to follow the evolution of any Web site
- Simple method
- Two main biases:
 - (1) age and permanence of Web locations and
 - (2) search engine dependence

Conclusion

- In the age of link-lead ranking algorithms and Google, there can be little doubt that the (absolute, external) WIF is an important measure of the impact of a Web space
- This experiment has drawn out issues that indicate it has a strong correlation with scientific production
- Webmasters need to take the WIF into account as a measure to attract visitors from search engines

Reading Factor

- Electronic access to full text journals has been available in university and hospital settings for a few years
- Specific software can be used to measure the numbers and types of articles consulted
- Based on the use of such software, we propose to develop a new means of evaluation of the usage of journals by end-users
- Because such an analysis directly relies upon the consultation of articles, we suggest to define a new bibliometric criterion called "**reading factor**" (RF).

Definition of the RF

$$RF = \frac{\text{consultations of a given journal}}{\text{consultations of a mean journal*}}$$

*calculated by dividing the total number of electronic accesses by the number of journals in the database

$$RF_j = \frac{C_j}{\sum C_j}$$

Example for Lancet

Number of clicks (Lancet) = $C_{\text{Lancet}} = 901$

Total number of electronic accesses = 8,280

Total of journals = 45

RF (Lancet) = $901 / (8280/45) = 4.90$

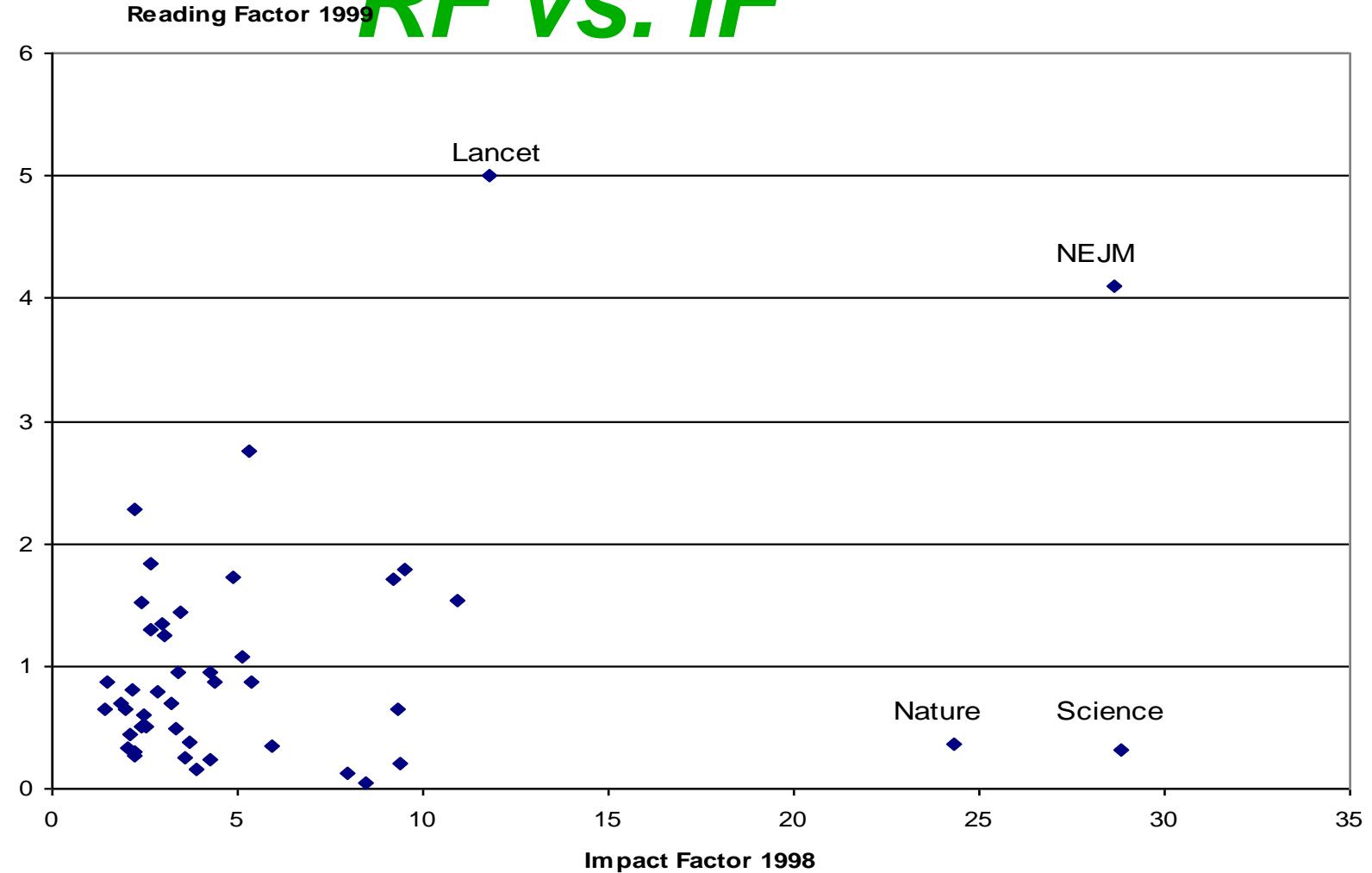


RF: Top 10

	RF 1999	IF 1998
Lancet	4.90	11.793
New England Journal of Medicine	4.01	28.66
JAMA	1.74	9.522
Journal of Urology	1.80	2.685
Circulation	1.67	9.173
Chest	2.23	2.246
British Medical Journal	2.69	5.325
British Journal of Surgery	1.49	2.381
Annals of Internal Medicine	1.50	10.900
British Journal of Surgery	1.49	2.381



RF vs. IF



RF vs. IF (cont.)

- Pearson's correlation coefficient between IF and RF was positive ($r=0.29$) and borderline significant ($p=0.052$)
- Distributions of IF and RF were clearly not normal => additional analyses
 - First, NEJM was the most influential journal in this analysis and strongly induced a positive correlation between IF and RF because of its very high IF (28.7) and RF values (4.1). Correlation coefficient calculated for 44 journals (excluding this journal) was no longer significant ($r=0.046$, $p=0.77$).
 - Second, no significant correlation could be found either using log transformed variables ($r=0.048$, $p=0.75$) or the Spearman rank correlation coefficient ($r=0.083$, $p=0.59$) on all 45 journals.

RF: a new bibliometric tool to manage a digital library

- Decision to change our Digital Library (DL) from Intranet to Internet in September 2000 because:
 - Much better update (>3 months)
 - Less work load in the Computing Department
 - BUT more expensive
- Necessity to diminish the number of journals
 - 45 >> 35
- Decision to choose the RF as the final criterion to manage the digital library collection... instead of a human consensus difficult to find
 - Empirical rule: if $RF < 0.5$ and $IF < 10$ then the journal is excluded from the RUH DL (e.g. American Journal of Psychiatry, American Journal of Public Health, Archives of Dermatology, Archives of General Psychiatry, Arteriosclerosis)

RF: a new bibliometric tool to manage a digital library (cont.)

- Because some other journals (e.g. British Medical Journal) were available freely on the Internet or with a password in a package with the paper version, the RUH DL had the opportunity to add some new journals spending the same amount of money (22,000 €/y).
- To choose **new journals** to include into our digital library in September 2000, we use the following indicators:
 - Brandon/Hill List of Journals
 - Abridged Index Medicus (AIM) Journal Titles (N=120) => Core Clinical Journal (PubMed)
 - *Journal Citation Reports*
 - Previous coverage in terms of medical specialities
- Results: Archives of otolaryngology-head and neck surgery, Arthritis & rheumatism, Clinical orthopaedics and related research, Clinical pharmacology and therapeutics, Critical care medicine
- Next year, RF will be applied to these new journals

Cost-analysis of the RUH DL

- Possible to use the total number of electronic accesses during a year for cost-analysis:
 - In 1998, 5,007; in 1999, 8,280; in 2000, 10,893
- An article in interlibrary loan costs an average of 4€ in France.
- As the RUH DL costs annually 22,000 €, cost of an electronic article was 4.13 € in 1998, 2.66 € in 1999 and 2.02 in 2000

=> RUH DL is cost effective since 1999

Discussion

- RF provides different bibliometric information than IF
- The frequency of use of a journal could therefore be a significant parameter of its interest to the readers and could be used as a more relevant marker of a given journal's influence.
- Electronic access allows a reliable and automatic analysis of the consultation rate, with a minimal delay, while conversely, IF is delayed by 2 to 4 years according to its calculation method
- RF provides an appreciation of the readers' interests in real time

Conclusion (cont.)

- A typology of readerships and journals could be achieved by a correlative examination of RF and IF.
- Biomedical publications could be classified in:
 - those of scientific interest ($IF > RF$),
 - those of major scientific interest ($IF >> RF$),
 - those of clinical interest ($RF > IF$),
 - those of major clinical interest ($RF >> IF$), and
 - those of mixed interest ($RF \approx IF$).

- L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle. Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit à la vente ou à la location. Sa diffusion, duplication, mise à disposition du public (sous quelque forme ou support que ce soit), mise en réseau, partielles ou totales, sont strictement réservées à l'université de Rouen.

L'utilisation de ce document est strictement réservée à l'usage privé des étudiants inscrits à l'UFR de médecine de l'université Rouen, ainsi que ceux inscrits au C2I Santé, et non destinée à une utilisation collective, gratuite ou payante.

Ce document a été réalisé par la Cellule TICE Médecine de la Faculté de Médecine de Rouen (Courriel : Francoise.Charles@univ-rouen.fr).