SUBJECT DELINEATION, CLASSIFICATION AND INFORMATION RETRIEVAL FOR BIBLIOMETRIC USE

Wolfgang Glänzel

Centre for R&D Monitoring and Dept. MSI, KU Leuven, Belgium

ECOOM

Structure of presentation

Classification of science into a disciplinary structure is at least as old as science itself. After many centuries of constructive but yet inconclusive search for a perfect classification scheme, the only sensible approach to the question appears to be the pragmatic one: what is the optimal scheme for a given practical purpose?

Narin, *Evaluative Scientometrics*, 1976

> *Classification of science into a disciplinary structure is at least as old as science itself. After many centuries of constructive but yet inconclusive search for a perfect classification scheme, the only sensible approach to the question appears to be the pragmatic one: what is the optimal scheme for a given practical purpose?*
>
> Narin, *Evaluative Scientometrics*, 1976

Subject classification and subject delineation has two important applications with different purposes

- Information services
- Research evaluation

Classification for standard applications

Classification systems have early been developed by the producer of the *Science Citation Index* (ISI), by institutions working extensively with this database and by the producers of other multidisciplinary journal databases [🗎 Narin, *Evaluative Scientometrics*, 1976].

Classification for standard applications

Classification systems have early been developed by the producer of the *Science Citation Index* (ISI), by institutions working extensively with this database and by the producers of other multidisciplinary journal databases [🗎 Narin, *Evaluative Scientometrics*, 1976].

Most are based on journal assignment, originally created for retrieval purposes, and have shortcomings when used in the context of research evaluation.

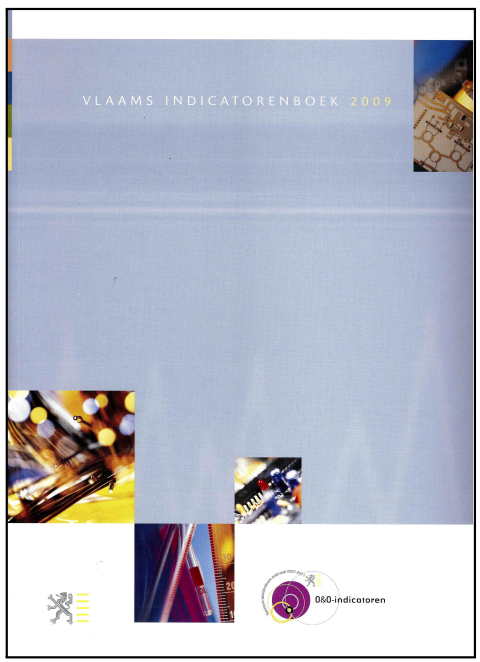
Classification for standard applications

Classification systems have early been developed by the producer of the *Science Citation Index* (ISI), by institutions working extensively with this database and by the producers of other multidisciplinary journal databases [🗎 Narin, *Evaluative Scientometrics*, 1976].

Most are based on journal assignment, originally created for retrieval purposes, and have shortcomings when used in the context of research evaluation.

Within bibliometrics there are standard applications which require stable hierarchically structured classification schemes. These are used as *general-purpose structures*.

Some examples



ISI Classification schemes

Two subject classification systems were introduced by ISI/Thomson Reuters.

1. ISI Subject Categories (part of the citation indexes and the JCR)
   - Fine grained (ca. 250 categories)
   - Forms a fuzzy system with multiple assignments
2. ESI Fields (part of the Essential Science Indicators)
   - Coarse classification (22 fields)
   - Forms a partition with unique assignment

☞ Elsevier's Scopus Subject Areas and Subject Categories system is based on journal classification with 4 areas, 27 major thematic categories and 313 specific subject categories.

The ECCOM classification scheme

I) The "cognitive" approach (setting the categories):
An initial scheme was elaborated on the basis of both the experience of bibliometricians and external experts.

The ECCOM classification scheme

I) The "cognitive" approach (setting the categories):
An initial scheme was elaborated on the basis of both the experience of bibliometricians and external experts.

II) The "pragmatic" approach (journal classification):
The journal set extracted from the WoS was classified into the preset subfields. The scheme has been adjusted according to co-heading frequency to keep multiple assignments within reasonable limits.

The ECCOM classification scheme

I) The "cognitive" approach (setting the categories):
An initial scheme was elaborated on the basis of both the experience of bibliometricians and external experts.

II) The "pragmatic" approach (journal classification):
The journal set extracted from the WoS was classified into the preset subfields. The scheme has been adjusted according to co-heading frequency to keep multiple assignments within reasonable limits.

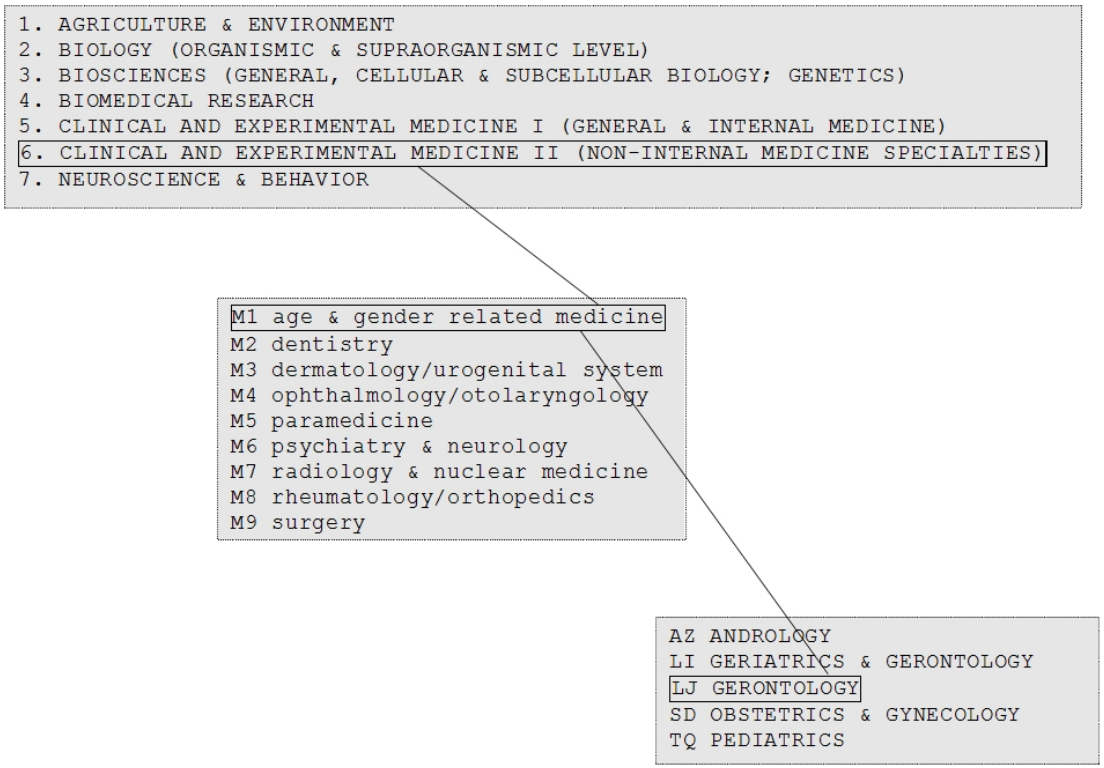
III) The "bibliometric" approach (article classification):
Articles published in core journals can be unambiguously classified into the subfield of the given journals. Articles of ambiguously assignable journals are classified individually using the analysis of references.

The ECCOM classification scheme

The hierarchical structure of the ECCOM scheme

0. level: 3 main areas
1. level: 15 major fields
2. level: 65 sub-fields
3. level: > 200 subject categories

Example of the sciences



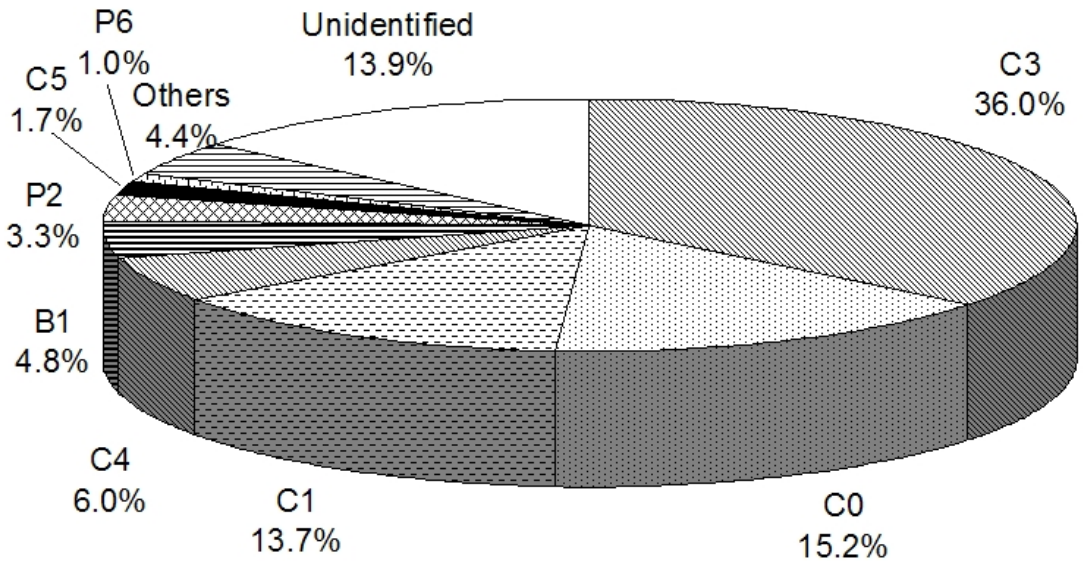
The ECCOM classification scheme

The subject categories after step II

1. **AGRICULTURE & ENVIRONMENT**
   - A1 Agricultural Science & Technology
   - A2 Plant & Soil Science & Technology
   - A3 Environmental Science & Technology
   - A4 Food & Animal Science & Technology
2. **BIOLOGY (ORGANISMIC & SUPRAORGANISMIC LEVEL)**
   - Z1 Animal Sciences
   - Z2 Aquatic Sciences
   - Z3 Microbiology
   - Z4 Plant Sciences
   - Z5 Pure & Applied Ecology
   - Z6 Veterinary Sciences
3. **BIOSCIENCES (GENERAL, CELLULAR & SUBCELLULAR BIOLOGY; GENETICS)**
   - B0 Multidisciplinary Biology
   - B1 Biochemistry/Biophysics/Molecular Biology
   - B2 Cell Biology
   - B3 Genetics & Developmental Biology
4. **BIOMEDICAL RESEARCH**
   - R1 Anatomy & Pathology
   - R2 Biomaterials & Bioengineering
   - R3 Experimental/Laboratory Medicine
   - R4 Pharmacology & Toxicology
   - R5 Physiology
5. **CLINICAL AND EXPERIMENTAL MEDICINE I (GENERAL & INTERNAL MEDICINE)**
   - I1 Cardiovascular & Respiratory Medicine
   - I2 Endocrinology & Metabolism
   - I3 General & Internal Medicine
   - I4 Hematology & Oncology
   - I5 Immunology
6. **CLINICAL AND EXPERIMENTAL MEDICINE II (NON-INTERNAL MEDICINE SPECIALTIES)**
   - M1 Age & Gender Related Medicine
   - M2 Dentistry
   - M3 Dermatology/Urogenital System
   - M4 Ophthalmology/Otolaryngology
   - M5 Paramedicine
   - M6 Psychiatry & Neurology
   - M7 Radiology & Nuclear Medicine
   - M8 Rheumatology/Orthopedics
   - M9 Surgery
7. **NEUROSCIENCE & BEHAVIOR**
   - N1 Neurosciences & Psychopharmacology
   - N2 Psychology & Behavioral Sciences
8. **CHEMISTRY**
   - C0 Multidisciplinary Chemistry
   - C1 Analytical, Inorganic & Nuclear Chemistry
   - C2 Applied Chemistry & Chemical Engineering
   - C3 Organic & Medicinal Chemistry
   - C4 Physical Chemistry
   - C5 Polymer Science
   - C6 Materials Science
9. **PHYSICS**
   - P0 Multidisciplinary Physics
   - P1 Applied Physics
   - P2 Atomic, Molecular & Chemical Physics
   - P3 Classical Physics
   - P4 Mathematical & Theoretical Physics
   - P5 Particle & Nuclear Physics
   - P6 Physics of Solids, Fluids And Plasmas
10. **GEOSCIENCES & SPACE SCIENCES**
    - G1 Astronomy & Astrophysics
    - G2 Geosciences & Technology
    - G3 Hydrology/Oceanography
    - G4 Meteorology/Atmospheric & Aerospace Science & Technology
    - G5 Mineralogy & Petrology
11. **ENGINEERING**
    - E1 Computer Science/Information Technology
    - E2 Electrical & Electronic Engineering
    - E3 Energy & Fuels
    - E4 General & Traditional Engineering
12. **MATHEMATICS**
    - H1 Applied Mathematics
    - H2 Pure Mathematics
13. **SOCIAL SCIENCES I (GENERAL, REGIONAL & COMMUNITY ISSUES)**
    - S1 Education & Information
    - S2 General, Regional & Community Issues
14. **SOCIAL SCIENCES II (ECONOMICAL & POLITICAL ISSUES)**
    - O1 Economics, Business & Management
    - O2 History, Politics & Law
15. **ARTS & HUMANITIES**
    - U1 Arts & Literature
    - U2 Language & Culture
    - U3 Philosophy & Religion

Journal classification after step III

[assignment of papers in 'Angewandte Chemie – International Edition' (1993)]



Subject delineation

Subject delimitation has become a central issue in so-called "domain studies" and the bibliometric studies of interdisciplinary research.

- Science policy addresses new emerging or complex interdisciplinary topics the delineation of which is particularly difficult.
  - The delineation of such domains is strongly related with information retrieval (e.g., using core journals, keywords and phrases),
  - but goals and methods of advanced subject delineation essentially differ from those of traditional retrieval.

Subject delineation

Subject delimitation has become a central issue in so-called "domain studies" and the bibliometric studies of interdisciplinary research.

- Science policy addresses new emerging or complex interdisciplinary topics the delineation of which is particularly difficult.
  - The delineation of such domains is strongly related with information retrieval (e.g., using core journals, keywords and phrases),
  - but goals and methods of advanced subject delineation essentially differ from those of traditional retrieval.

☞ Proper subject delineation is also necessary to find correct reference standards for benchmarking the research performance of the actors in the topic under study.

Why subject delineation?

Sufficiently fine-grained intellectual subject classification schemes can help define a broader scope within the actual subject has to be delineated.

However, using preset disciplines or topics usually results in noise that is too large for obtaining acceptable coverage with both high precision and recall.

Even scientific journals are too coarse for subject delineation since the distribution of relevant documents over journals is very skewed (cf. *Bradford's Law*).

Main fields of applications:

- Interdisciplinary subjects
  - Literature is spread over a plethora of specific, general and multidisciplinary journals.
- New emerging topics
  - Literature is spread over various journals since own specific communication channels might often not yet exist.
- Activity of institutes and research groups
  - Institutes might have a special focus and subject delineation might thus be influenced by the institutional profile.

Bradford's Law

In 1934 Samuel C. Bradford published a study on the frequency distribution of papers over journals. He found that

> *"If scientific journals are arranged in order of decreasing productivity on a given subject, they may be divided into a nucleus of journals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus when the numbers of periodicals in the nucleus and the succeeding zones will be as* $1 : b : b^2 \ldots$*"*

Bradford, *Engineering*, 1934

Why subject delineation?

Bradford plot for Tetrachloro-dibenzo-dioxin literature in PubMed between 1976 and 2005 according to Peña-Rey et al. (2006)

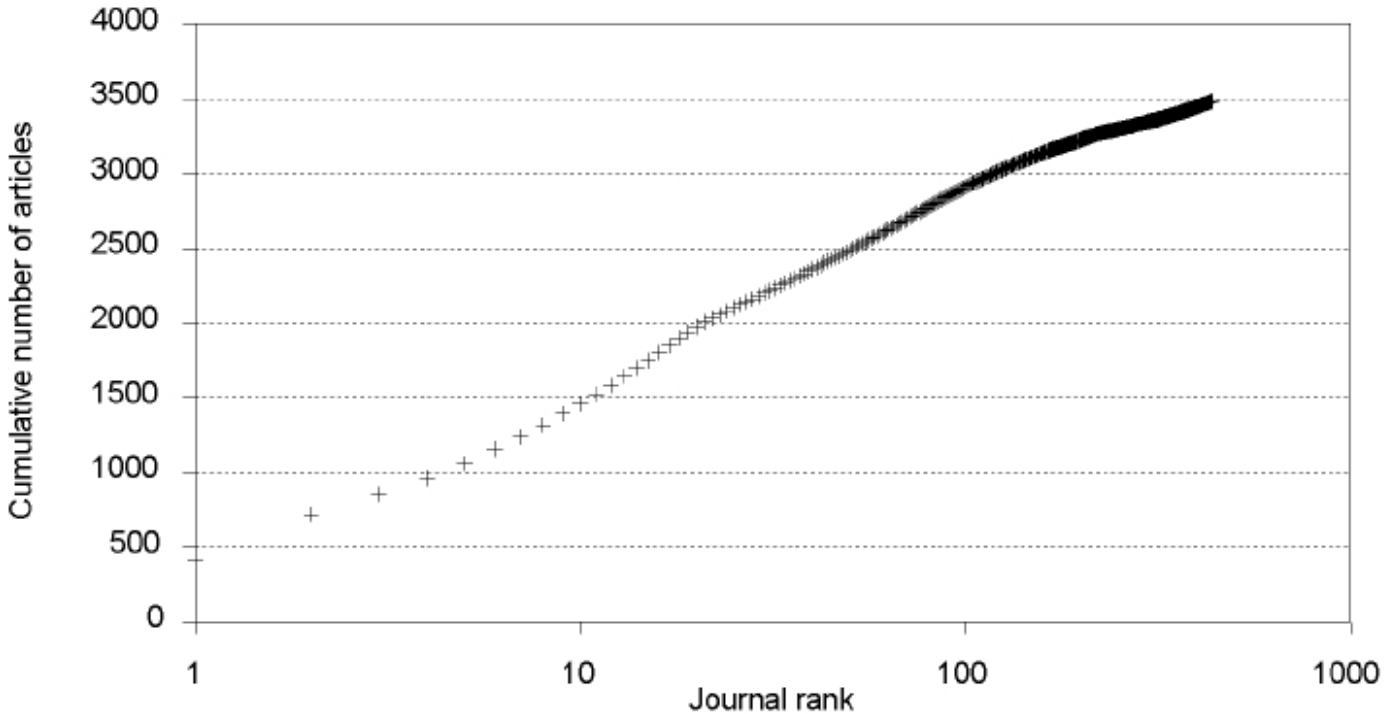


Figure redrawn from Peña-Rey et al., *Scientometrics,* 2006

Why subject delineation?

Bradford data on TCDD literature in the period 1976-2005

Zone	Journals	Articles	b
[nucleus]	1	404	—
Z1	2	450	2.0
Z2	5	463	2.5
Z3	7	432	1.4
Z4	12	393	1.7
Z5	31	433	2.6
Z6	57	412	1.8
Z7	321	497	5.6

Source: Peña-Rey et al., *Scientometrics*, 2006

Information Retrieval for Bibliometrics?

Application of bibliometrics has increasingly shifted towards meso and micro studies in the sense of both actors and topic analysis.
One consequence is the necessity of proper subject delineation (domain studies, interdisciplinary research, emerging topics).
Subject delineation strongly relies on IR methods through complex *search strategies*.

Bibliometrics for Information Retrieval?

Bibliometrics, in turn, provides important techniques to improve the efficiency of IR. Similarity/distance measures defined on direct citations, bibliographic coupling, lexical relationship and or even "core documents" can facilitate and improve the retrieval of scientific information.

Bibliometrics for Information Retrieval?

Bibliometrics, in turn, provides important techniques to improve the efficiency of IR. Similarity/distance measures defined on direct citations, bibliographic coupling, lexical relationship and or even "core documents" can facilitate and improve the retrieval of scientific information.

Both Bibliometrics and Information Retrieval may serve as mutual input and can be combined in an iterative way. This combination will be shown in the second part of the presentation.

☞ The first part of the lecture, however, will focus on traditional techniques that can easily be developed for and used in the online versions of bibliographic databases.

Traditional delineation

Traditional retrieval is usually a combination of core journals and keywords, phrases and terms. In addition corporate addresses and author names can be used.

Advantages:

- This type is user-friendly: Search strategies can readily be built and applied.
- Modifications for following the evolution of the subject in study are possible and relatively easy.

Disadvantages:

- This type tends to produce “wild shoots”. The strategy might mushroom and grow very complex and too specific.
- The effect of adding/removing terms/phrases might be beyond control.
- Terms might turn redundant. Adding new terms does not essentially increase the number of hits.
- The effect of logically combining terms is no longer transparent.
- The evolution of the subject cannot longer be captured by moderate modifications of the strategy.

Traditional delineation

Disadvantages:

- This type tends to produce “wild shoots”. The strategy might mushroom and grow very complex and too specific.
- The effect of adding/removing terms/phrases might be beyond control.
- Terms might turn redundant. Adding new terms does not essentially increase the number of hits.
- The effect of logically combining terms is no longer transparent.
- The evolution of the subject cannot longer be captured by moderate modifications of the strategy.

☛ Avoid too complex strategies that produce unpredictable results and are not flexible enough for dynamic analysis.

Traditional delineation

The delineation of the research field "wine research" based on traditional retrieval (Glänzel & Veugelers, 2006)

The search strategy combined three components:

(1) Keywords, title and abstract:
GRAPEVIN* OR WINES OR WINE GRAP* OR WINE PRO* OR RED WINE* OR WHITE WINE* OR WINEMAKING OR ENOLOG* OR VITICULT* OR OENOLOG* OR WINE CELL* OR WINE YEAST* OR WINERY OR WINERIES

(2) Corporate address:
VITICULT* OR [O]ENOL*

(3) Cores journals:
AMERICAN JOURNAL OF ENOLOGY AND VITICULTURE
AUSTRALIAN JOURNAL OF GRAPE AND WINE RESEARCH

Example of a complex strategy

Nanotechnology

Search terms:

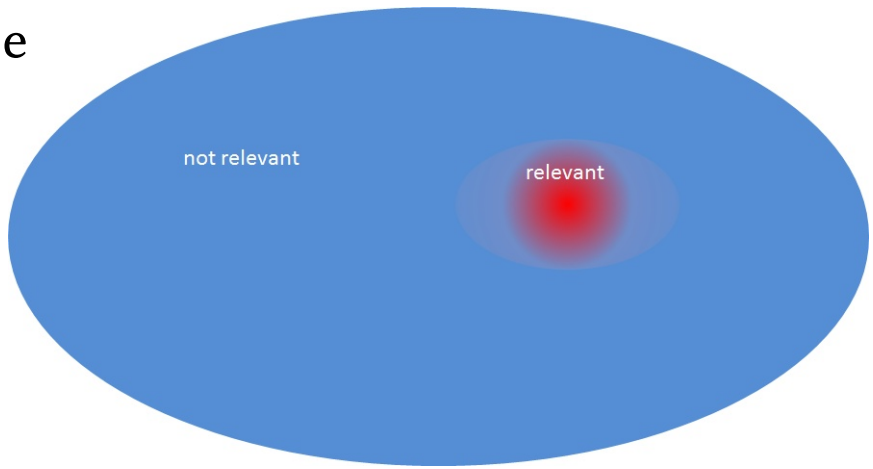
nano* NOT (nanomet* OR nano2 OR nano3 OR nano4 OR nano5 OR nanosecon* OR nano secon*)
OR
nanomet* scale* OR nanometerscale* OR nanometer length OR nano meter length
nanoa* OR nanob* OR nanoc* OR nanod* OR nanoe* OR nanof* OR nanog* OR nanoh* OR nanoi OR nanoj* OR nanok* OR nanol* OR nanon* OR nanoo* OR nanop* OR nanoq* OR nanor* OR nanot* OR nanou* OR nanov* OR nanow* OR nanox* OR nanoy* OR nano z* OR nano
atom* force microscop*
tunnel* microscop*
scanning probe microscop*
scanning force microscop*
semiconductor quantum dot
silicon quantum dot
quantum dot array
coulomb blockade
self-organized growth
Drug carriers
positional assembly
modified virus
molecular templates
supramolecular chemistry
(drug delivery OR drug targeting OR gene therapy OR gene delivery) AND (polymer OR particles OR encapsulation OR conjugate)
Immobilized AND (DNA OR template OR primer OR oligonucleotide OR polynucleotide)
Polymer AND (protein OR antibody OR enzyme OR DNA OR RNA OR polynucleotide OR virus)
Surface modification AND (self assembling OR molecular layers OR multilayer OR layer-by-layer)
Self assembling AND (biocompatibility OR bloodcompatibility OR blood compatibility OR cellseeding OR cell seeding OR cell therapy OR tissue repair OR extracellular matrix OR tissue engineering OR biosensors OR immunosensor OR biochip OR nano-particles OR cell adhesion)
Site-specific AND (gene therapy OR drug delivery OR gene delivery)
Encapsulation AND virus
(Patterns OR patterning) AND (organized assemblies OR biocompatibility OR bloodcompatibility OR blood compatibility OR cellseeding OR cell seeding OR cell therapy OR tissue repair OR extracellular matrix OR tissue engineering OR biosensors OR immunosensor OR biochip OR cell adhesion)
Single molecule
molecular motor
molecular beacon
biosensor

Noyons et al., *Mapping Excellence in Science and Technology across Europe*, 2003

The initial situation

Bibliometrics, in general, requires specific retrieval. The borderline between relevant and not relevant documents is fuzzy and often determined by users or the actors in the domain in question. Sometimes it has to be adjusted according to the actual needs.

The scope of the study decides whether documents in the red or the purple circle are used for the bibliometric analysis.



The background

- The objectives of subject delineation in the framework of domain studies essentially differ from the goals of traditional information retrieval.
- In addition, bibliometrics allows including also 'metric' components in the search strategy.
- Thresholds of the strength of citation, bibliographic-coupling or textual links can be used to fine-tune the metric component.

Bibliometrics-aided retrieval is a combination of traditional search strategies with advanced bibliometrics methods.

- Zitt & Bassecaulard, *IPM*, 2006
- Glänzel et al., *STI Conference*, 2006

Precision and Recall

Two measures are defined to estimate the performance of a retrieval algorithm.

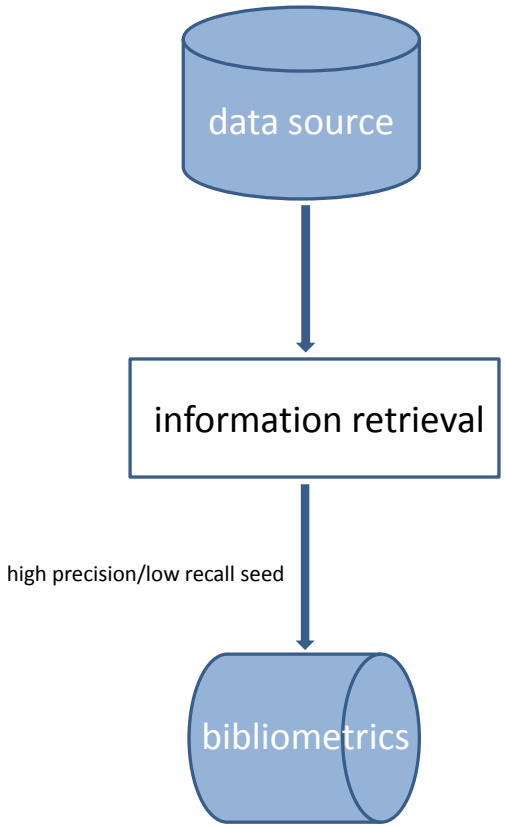
- Precision refers to the fraction of relevant documents in the retrieved set. It reflects the absence of type I errors (false positives). The precision can readily be calculated.
- Recall refers to the fraction of retrieved relevant documents in respect to the total number of relevant documents in the database. It reflects the absences of type II errors (false negatives). It is rather difficult to estimate the recall as the total number of relevant documents is usually unknown.

Bibliometrics-aided retrieval

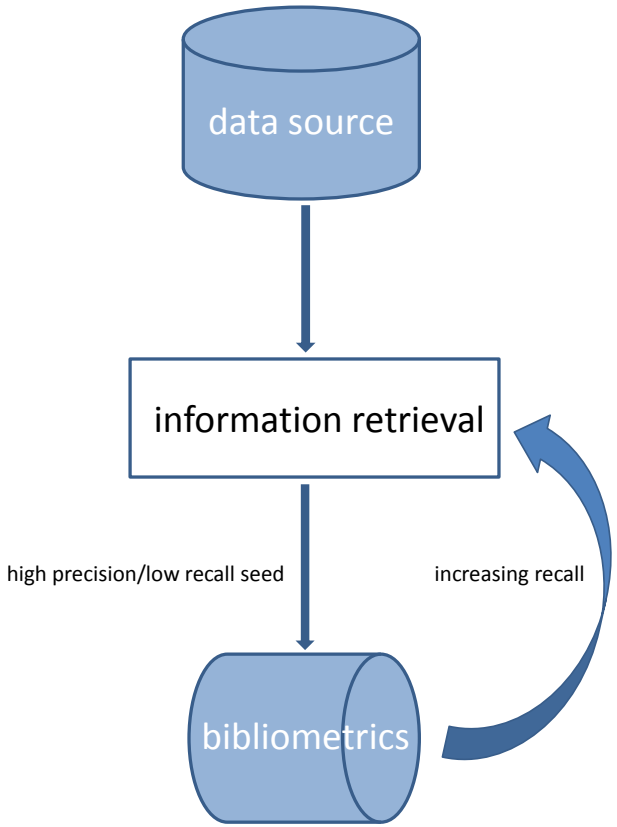
The aim of applying bibliometric techniques to IR is to develop retrieval strategies with high scores on both precision and recall, since otherwise increasing precision often results in a lower recall.

- ☞ Depending on the goal of the retrieval exercise, the importance can be shifted between these measures.

How to combine Information Retrieval and bibliometrics?



How to combine Information Retrieval and bibliometrics?



When Bibliometrics meets Information Retrieval ...

A method of enhancing subject delineation without inflating lexical queries by defining huge sequences of logical term/phrase combinations was proposed almost simultaneously by Zitt and Bassecoulard (2006) and Glänzel et al. (2004, 2006).

The method "combines a high-precision and low-recall seed, obtained by journal and lexical queries, and a citation-based extension enhancing the recall".

Laures et al., *Scientometrics*, 2010

Glänzel et al. called the seed *core set*.

When Bibliometrics meets Information Retrieval ...

The first step comprises so-called *unconditional* criteria: $(UC_1, \ldots, UC_k)$ with $k \geq 0$.

The resulting core set contains the seed documents.

In a second step this set is extended by potentially relevant documents. This includes so-called *conditional* criteria $\{CC_1, \ldots, CC_m, \ldots, CC_{m+n}\}$ with $m, n > 0$ or $m = n = 0$. Assume that we have at least one conditional criterion and another conditional or unconditional criterion.

The *bibliometric retrieval* (BR) strategy can then be defined as the following logical combination

$$BR = (UC_1 \vee \ldots \vee UC_k) \vee ((CC_1 \vee \ldots \vee CC_m) \wedge (CC_{m+1} \vee \ldots \vee CC_{m+n})).$$

Example: Stem cells (e.g., Glänzel et al., 2004)

UC1: Journal in WoS = STEM CELLS

UC2: Address word = STEM CELL*

UC3: Keywords = (STEM CELL* OR STEM (ES) CELL* OR PROGENITOR* CELL* OR HEMATOPOI* CELL*)

CC1: Journal = JOURNAL OF HEMATOTHERAPY & STEM CELL RESEARCH

CC2: Keywords = (BONE-MARROW OR UMBILICAL-CORD-BLOOD OR UCB OR HUCB OR CYTOPOI* OR MEGAKARYOPOI* OR ERYTHROPOI* OR MYELOPOI* OR THROMBOPOI* OR STROMAL CELL* OR PRECURSOR CELL*)

CC3: Cited source[1] = UC1 OR UC2 OR UC3

The search strategy: $BR := (UC1 \vee UC2 \vee UC3) \vee ((CC1 \vee CC2) \wedge CC3)$

[1] Papers citing 3–5 other papers classified as unconditionally relevant making up at least 40% of all SCIE references, or 6–10 UC papers making up at least 30% of all SCIE references, or citing more than 10 UC papers.

Example: Bioinformatics (e.g., Glänzel et al., 2006, 2009)

UC1: Journal in WoS = BIOINFORMATICS (formerly COMPUTER APPLICATIONS IN THE BIOSCIENCES), JOURNAL OF COMPUTATIONAL BIOLOGY, BRIEFINGS IN BIOINFORMATICS, BMC BIOINFORMATICS

UC2: Journal in Medline= IN SILICO BIOLOGY, PSB ON-LINE PROCEEDINGS, APPLIED BIOINFORMATICS, PLOS COMPUTATIONAL BIOLOGY

CC1: Keywords in title = BIOINFORMATICS, COMPUTATIONAL BIOLOG*, SYSTEMS BIOLOGY

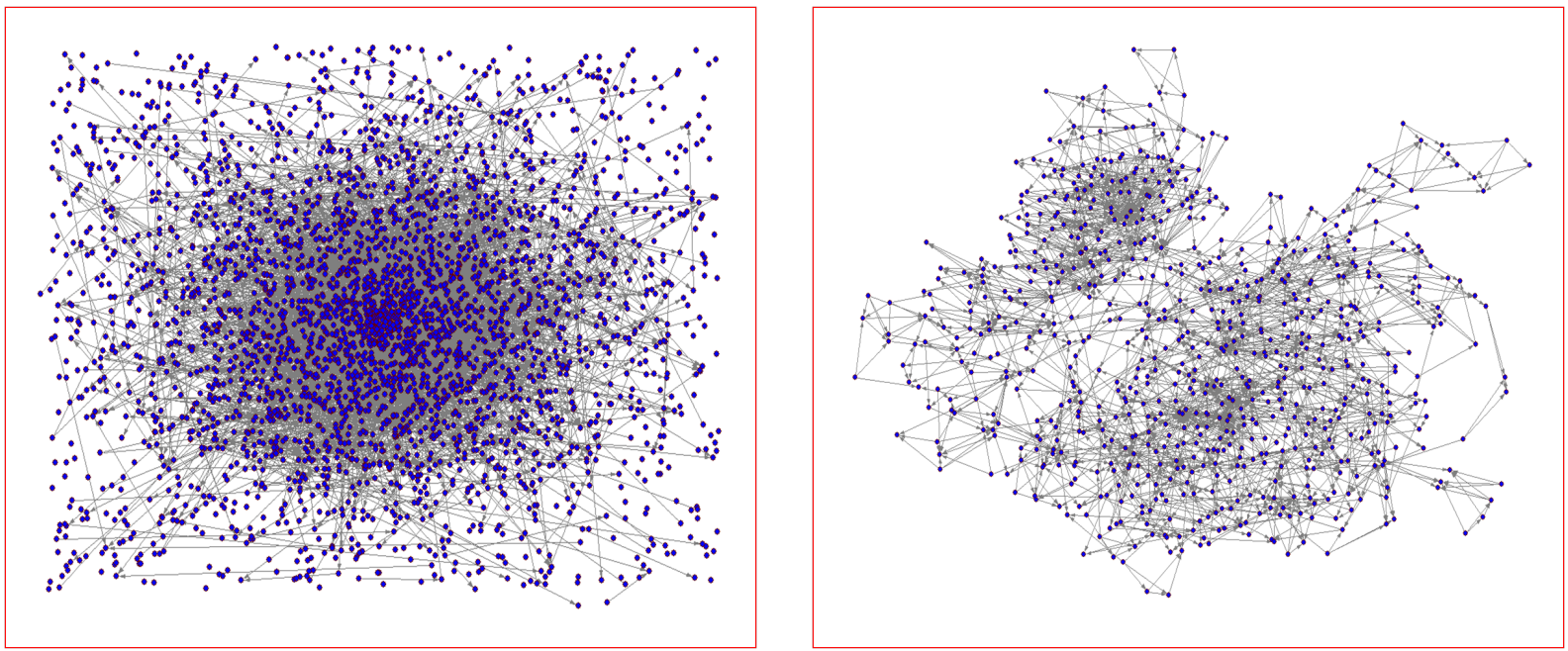
CC2: Related records of UC1

CC3: Cited or citing source[1] of UC1

The search strategy: $BR := (UC_1 \vee UC_2) \vee (CC_1 \wedge (CC_2 \vee CC_3))$

[1] Different rules for citations (both directions) can be defined.

The core set (UC_1 – left) and of all retrieved Bioinformatics papers (BR – right) with Kamada-Kawai layout (Glänzel et al., 2006)



Data source: Thomson Reuters Web of Knowledge

Bibliographic coupling as a measure of relatedness

- Instead of direct citations (one or both directions) bibliographic coupling can be used in bibliometrics-aided retrieval.
- This option is provided in the online versions of Thomson Reuters' Web of Science and Elsevier's Scopus database.
- Proper thresholds are needed to fine-tune the "distance" of potentially relevant documents from the "seed" ("core set").

Bibliographic coupling as a measure of relatedness

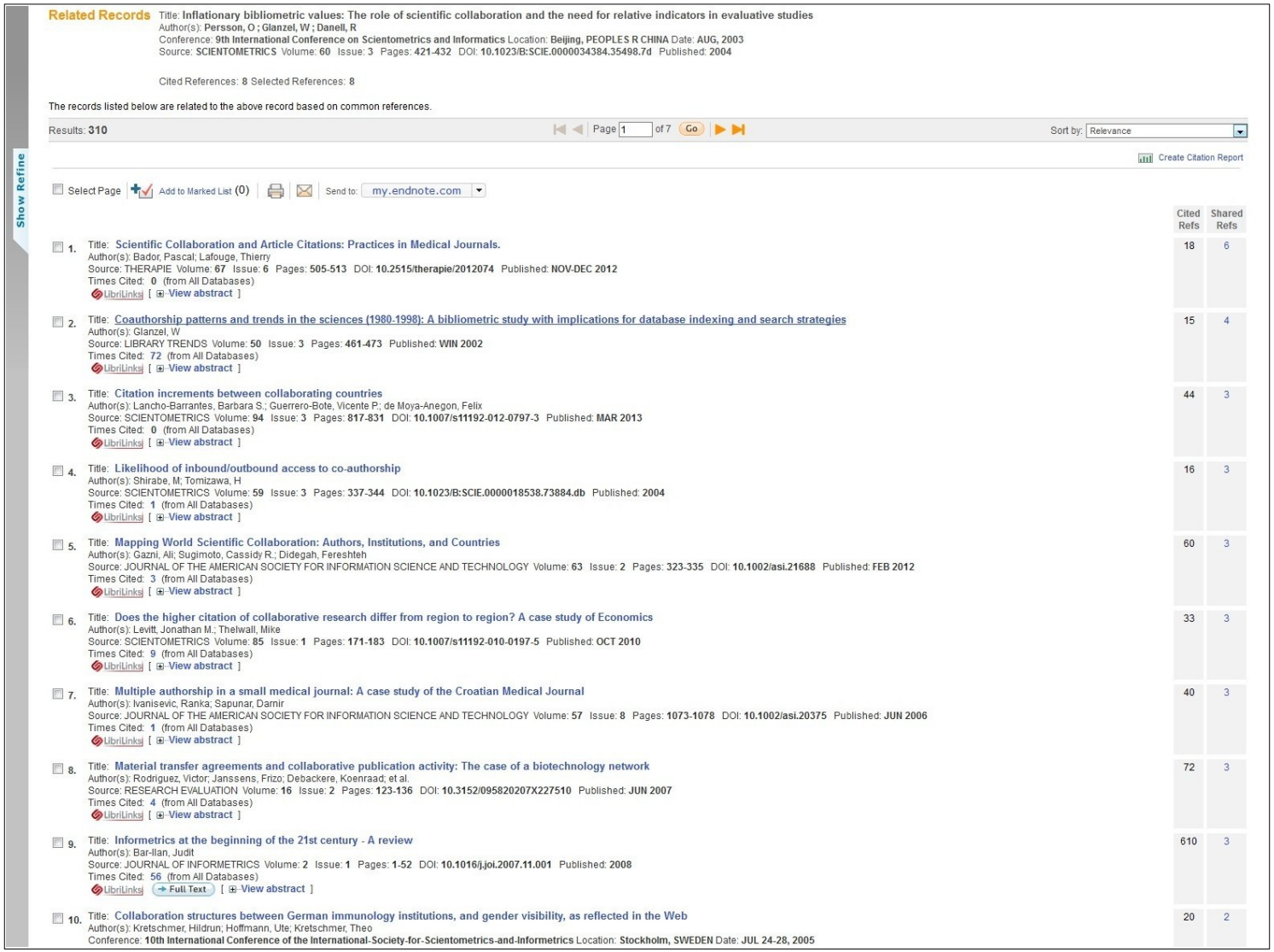
- The notion of *bibliographic coupling* has been introduced by Fano (1956) and Kessler (1963).
- According to this notion, documents are related if they share a considerable part of their reference lists.
- Salton's (cosine) measure as an indicator of the strength of relatedness:

$$r_{ij} = \frac{p_{ij}}{\sqrt{p_i \cdot p_j}},$$

where p_{ij} is the number of joint references and $p_i(p_j)$ the number of all references of the two documents i and j, respectively.

☞ The set-based *Jaccard Index* $J_{ij} = \frac{p_{ij}}{p_i + p_j - p_{ij}}$ can be used as an alternative measure.

Related records in the Web of Science database



Related Records Title: **Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies**
Author(s): **Persson, O ; Glanzel, W ; Danell, R**
Conference: **9th International Conference on Scientometrics and Informatics** Location: **Beijing, PEOPLES R CHINA** Date: **AUG, 2003**
Source: **SCIENTOMETRICS** Volume: **60** Issue: **3** Pages: **421-432** DOI: **10.1023/B:SCIE.0000034384.35498.7d** Published: **2004**

Cited References: 8 Selected References: 8

The records listed below are related to the above record based on common references.

Results: **310** Page 1 of 7 Go Sort by: Relevance

Create Citation Report

Show Refine

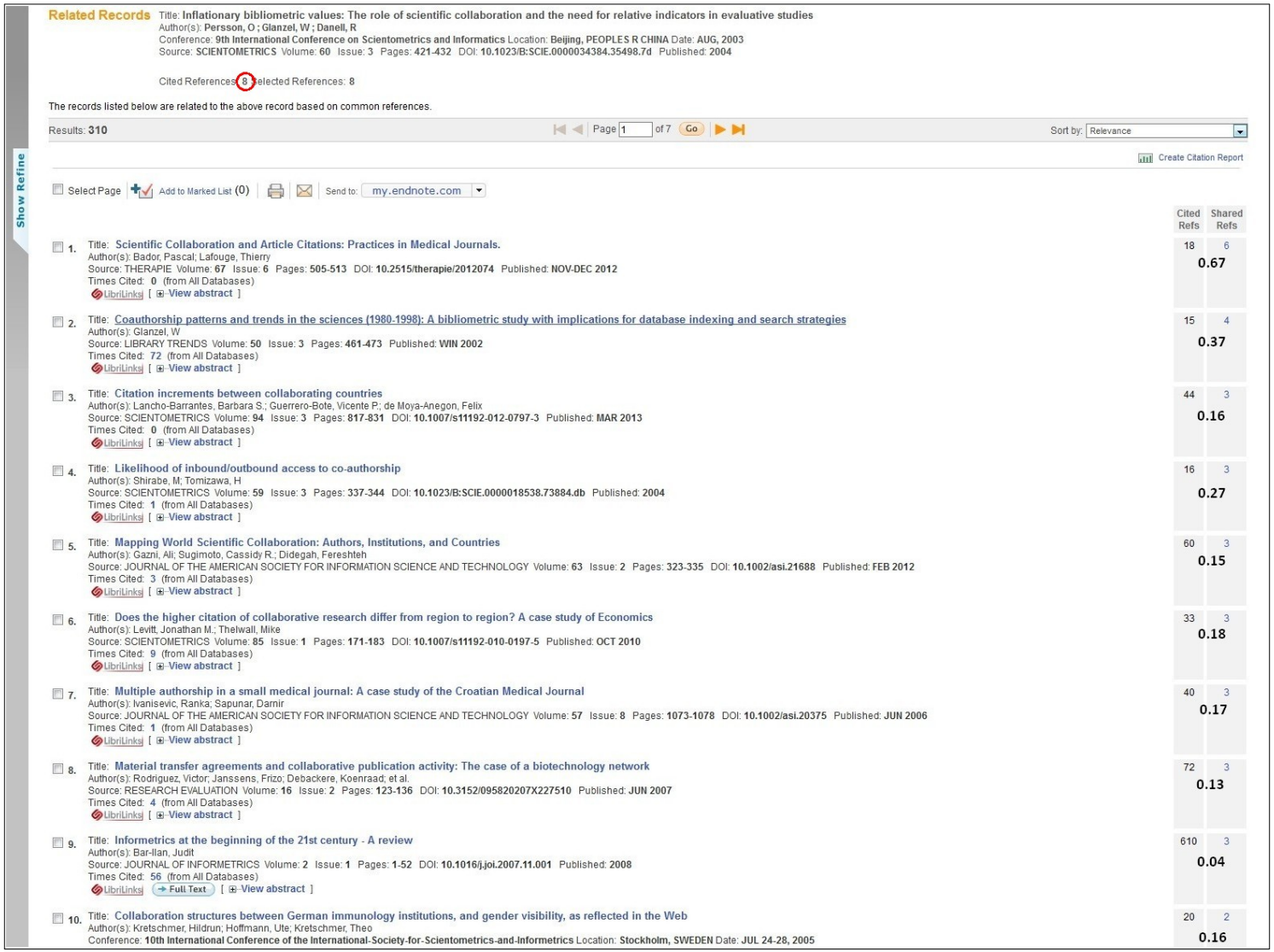
Select Page | Add to Marked List (0) | Send to: my.endnote.com

		Cited Refs	Shared Refs
1.	Title: **Scientific Collaboration and Article Citations: Practices in Medical Journals.** Author(s): Bador, Pascal; Lafouge, Thierry Source: THERAPIE Volume: **67** Issue: **6** Pages: **505-513** DOI: **10.2515/therapie/2012074** Published: **NOV-DEC 2012** Times Cited: **0** (from All Databases) LibriLinks [View abstract]	18	6
2.	Title: **Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies** Author(s): Glanzel, W Source: LIBRARY TRENDS Volume: **50** Issue: **3** Pages: **461-473** Published: **WIN 2002** Times Cited: **72** (from All Databases) LibriLinks [View abstract]	15	4
3.	Title: **Citation increments between collaborating countries** Author(s): Lancho-Barrantes, Barbara S.; Guerrero-Bote, Vicente P.; de Moya-Anegon, Felix Source: SCIENTOMETRICS Volume: **94** Issue: **3** Pages: **817-831** DOI: **10.1007/s11192-012-0797-3** Published: **MAR 2013** Times Cited: **0** (from All Databases) LibriLinks [View abstract]	44	3
4.	Title: **Likelihood of inbound/outbound access to co-authorship** Author(s): Shirabe, M; Tomizawa, H Source: SCIENTOMETRICS Volume: **59** Issue: **3** Pages: **337-344** DOI: **10.1023/B:SCIE.0000018538.73884.db** Published: **2004** Times Cited: **1** (from All Databases) LibriLinks [View abstract]	16	3
5.	Title: **Mapping World Scientific Collaboration: Authors, Institutions, and Countries** Author(s): Gazni, Ali; Sugimoto, Cassidy R.; Didegah, Fereshteh Source: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volume: **63** Issue: **2** Pages: **323-335** DOI: **10.1002/asi.21688** Published: **FEB 2012** Times Cited: **3** (from All Databases) LibriLinks [View abstract]	60	3
6.	Title: **Does the higher citation of collaborative research differ from region to region? A case study of Economics** Author(s): Levitt, Jonathan M.; Thelwall, Mike Source: SCIENTOMETRICS Volume: **85** Issue: **1** Pages: **171-183** DOI: **10.1007/s11192-010-0197-5** Published: **OCT 2010** Times Cited: **9** (from All Databases) LibriLinks [View abstract]	33	3
7.	Title: **Multiple authorship in a small medical journal: A case study of the Croatian Medical Journal** Author(s): Ivanisevic, Ranka; Sapunar, Damir Source: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volume: **57** Issue: **8** Pages: **1073-1078** DOI: **10.1002/asi.20375** Published: **JUN 2006** Times Cited: **1** (from All Databases) LibriLinks [View abstract]	40	3
8.	Title: **Material transfer agreements and collaborative publication activity: The case of a biotechnology network** Author(s): Rodriguez, Victor; Janssens, Frizo; Debackere, Koenraad; et al. Source: RESEARCH EVALUATION Volume: **16** Issue: **2** Pages: **123-136** DOI: **10.3152/095820207X227510** Published: **JUN 2007** Times Cited: **4** (from All Databases) LibriLinks [View abstract]	72	3
9.	Title: **Informetrics at the beginning of the 21st century - A review** Author(s): Bar-Ilan, Judit Source: JOURNAL OF INFORMETRICS Volume: **2** Issue: **1** Pages: **1-52** DOI: **10.1016/j.joi.2007.11.001** Published: **2008** Times Cited: **56** (from All Databases) LibriLinks Full Text [View abstract]	610	3
10.	Title: **Collaboration structures between German immunology institutions, and gender visibility, as reflected in the Web** Author(s): Kretschmer, Hildrun; Hoffmann, Ute; Kretschmer, Theo Conference: **10th International Conference of the International-Society-for-Scientometrics-and-Informetrics** Location: **Stockholm, SWEDEN** Date: **JUL 24-28, 2005**	20	2

Source: Thomson Reuters Web of Knowledge

Bibliometrics-aided retrieval

Related records in the Web of Science database



Related Records Title: **Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies**
Author(s): **Persson, O ; Glanzel, W ; Danell, R**
Conference: **9th International Conference on Scientometrics and Informatics** Location: **Beijing, PEOPLES R CHINA** Date: **AUG, 2003**
Source: **SCIENTOMETRICS** Volume: **60** Issue: **3** Pages: **421-432** DOI: **10.1023/B:SCIE.0000034384.35498.7d** Published: **2004**

Cited References: **8** Selected References: **8**

The records listed below are related to the above record based on common references.

Results: **310** Page 1 of 7 Go Sort by: Relevance

Create Citation Report

Show Refine

Select Page | Add to Marked List (0) | Send to: my.endnote.com

		Cited Refs	Shared Refs
1.	Title: **Scientific Collaboration and Article Citations: Practices in Medical Journals.** Author(s): Bador, Pascal; Lafouge, Thierry Source: THERAPIE Volume: **67** Issue: **6** Pages: **505-513** DOI: **10.2515/therapie/2012074** Published: **NOV-DEC 2012** Times Cited: **0** (from All Databases) LibriLinks [View abstract]	18	6 **0.67**
2.	Title: **Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies** Author(s): Glanzel, W Source: LIBRARY TRENDS Volume: **50** Issue: **3** Pages: **461-473** Published: **WIN 2002** Times Cited: **72** (from All Databases) LibriLinks [View abstract]	15	4 **0.37**
3.	Title: **Citation increments between collaborating countries** Author(s): Lancho-Barrantes, Barbara S.; Guerrero-Bote, Vicente P.; de Moya-Anegon, Felix Source: SCIENTOMETRICS Volume: **94** Issue: **3** Pages: **817-831** DOI: **10.1007/s11192-012-0797-3** Published: **MAR 2013** Times Cited: **0** (from All Databases) LibriLinks [View abstract]	44	3 **0.16**
4.	Title: **Likelihood of inbound/outbound access to co-authorship** Author(s): Shirabe, M; Tomizawa, H Source: SCIENTOMETRICS Volume: **59** Issue: **3** Pages: **337-344** DOI: **10.1023/B:SCIE.0000018538.73884.db** Published: **2004** Times Cited: **1** (from All Databases) LibriLinks [View abstract]	16	3 **0.27**
5.	Title: **Mapping World Scientific Collaboration: Authors, Institutions, and Countries** Author(s): Gazni, Ali; Sugimoto, Cassidy R.; Didegah, Fereshteh Source: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volume: **63** Issue: **2** Pages: **323-335** DOI: **10.1002/asi.21688** Published: **FEB 2012** Times Cited: **3** (from All Databases) LibriLinks [View abstract]	60	3 **0.15**
6.	Title: **Does the higher citation of collaborative research differ from region to region? A case study of Economics** Author(s): Levitt, Jonathan M.; Thelwall, Mike Source: SCIENTOMETRICS Volume: **85** Issue: **1** Pages: **171-183** DOI: **10.1007/s11192-010-0197-5** Published: **OCT 2010** Times Cited: **9** (from All Databases) LibriLinks [View abstract]	33	3 **0.18**
7.	Title: **Multiple authorship in a small medical journal: A case study of the Croatian Medical Journal** Author(s): Ivanisevic, Ranka; Sapunar, Damir Source: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volume: **57** Issue: **8** Pages: **1073-1078** DOI: **10.1002/asi.20375** Published: **JUN 2006** Times Cited: **1** (from All Databases) LibriLinks [View abstract]	40	3 **0.17**
8.	Title: **Material transfer agreements and collaborative publication activity: The case of a biotechnology network** Author(s): Rodriguez, Victor; Janssens, Frizo; Debackere, Koenraad; et al. Source: RESEARCH EVALUATION Volume: **16** Issue: **2** Pages: **123-136** DOI: **10.3152/095820207X227510** Published: **JUN 2007** Times Cited: **4** (from All Databases) LibriLinks [View abstract]	72	3 **0.13**
9.	Title: **Informetrics at the beginning of the 21st century - A review** Author(s): Bar-Ilan, Judit Source: JOURNAL OF INFORMETRICS Volume: **2** Issue: **1** Pages: **1-52** DOI: **10.1016/j.joi.2007.11.001** Published: **2008** Times Cited: **56** (from All Databases) LibriLinks Full Text [View abstract]	610	3 **0.04**
10.	Title: **Collaboration structures between German immunology institutions, and gender visibility, as reflected in the Web** Author(s): Kretschmer, Hildrun; Hoffmann, Ute; Kretschmer, Theo Conference: **10th International Conference of the International-Society-for-Scientometrics-and-Informetrics** Location: **Stockholm, SWEDEN** Date: **JUL 24-28, 2005**	20	2 **0.16**

Source: Thomson Reuters Web of Knowledge

In order to facilitate the retrieval, especially within rather small areas, BR can be extended by using hybrid similarities where bibliographic coupling is combined with lexical similarity.

Instead of the combination of citation links or "related records" (based on bibliographic coupling), similarities based on hybrid textual-citation methods can be applied to some of the conditional criteria. This might help avoid too many steps in the logical BR algorithm.

Example

A document is considered relevant if it meets some conditional criterion (CC_j), and is strongly linked based on a hybrid similarity measure to at least a certain number of documents meeting an unconditional criterion (UC_i).

Core documents for topic representation and retrieval

The notion of a "core" of literature goes back to *co-citation analysis*.

- Small, *JASIS*, 1973.

Definition:

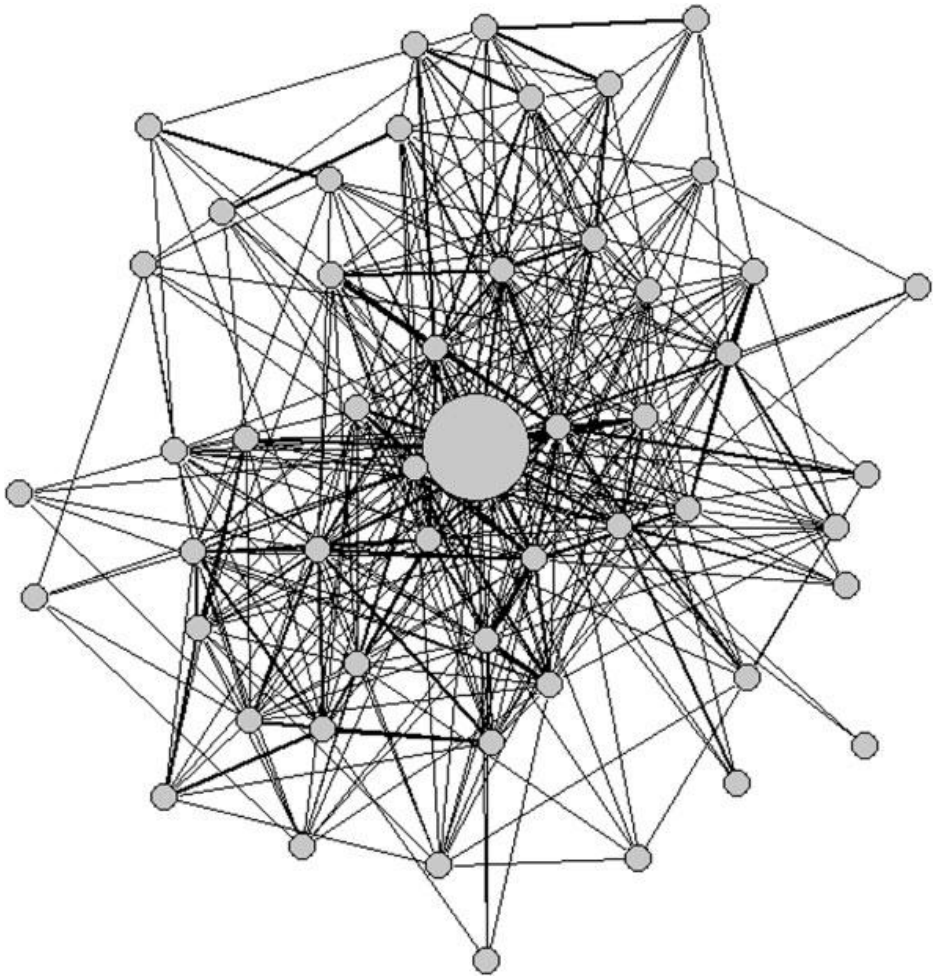
Core documents are defined as papers, which have at least *n* links of at least a given strength *r* according to a given similarity measure.

- Glänzel & Czerwon, *Scientometrics*, 1996
- Glänzel & Thijs, *Scientometrics*, 2012

Core documents can directly be used for document retrieval, namely to identify further relevant documents by following their strong and medium-strong links.

Cluster representation for dynamic analysis

Visualisation of the link environment of a 'core document'
(according to Glänzel & Thijs, 2012)



Data source: Thomson Reuters Web of Knowledge

Conclusions

The fields of applications of the described methods are manifold.
Below we give some examples.

- Bibliometrics-aided retrieval is a powerful tool to develop and adjust search strategy at any level of aggregation. It improves even the delineation of complex and interdisciplinary fields and topics.
- Metrics can be used for fine-tuning search strategies and to stop retrieval at any level.
- Bibliometrics-aided retrieval can thus be applied in dynamic analysis and if the scope of the subject is changed.
- Core documents represent the most interlinked papers in a set. Following their strong *and* weaker links might help retrieve relevant information without formulating search queries.
- Adjustable hybrid (text/citation-based) techniques allow bibliometrics-aided retrieval even in fields where citations do not play an important role.

Acknowledgement

Parts of this lecture are based on a joint presentation with Bart Thijs (ECOOM, KU Leuven) prepared for the *European Summer School for Scientometrics* (esss).

Thank you very much for your attention.

Vielen Dank für Ihre Aufmerksamkeit!

Hartelijk dank voor uw aandacht!

¡Muchísimas gracias por su atención!

Köszönöm szépen a figyelmüket!

Molte grazie per la vostra attenzione.

Muito obrigado pela vossa atenção.