

V.S.S.T. '2004

VEILLE STRATEGIQUE

SCIENTIFIQUE

TECHNOLOGIQUE

**SYSTÈMES D'INFORMATION ELABORÉE, BIBLIOMÉTRIE, LINGUISTIQUE,
INTELLIGENCE ÉCONOMIQUE**



TOULOUSE

25-29 octobre 2004

Organisation : IRIT associé à l'UPC et la SFBA

IRIT
Institut de Recherche en Informatique de Toulouse

V.S.S.T.'2004

Tome 2 : Salle B

Veille Stratégique Scientifique & Technologique

Systemes s'information élaborée

Bibliométrie, Linguistique,

Intelligence Economique

Toulouse

25-29 octobre 2004

Comité scientifique 2004

Président:

Bernard DOUSSET, IRIT-SIG, Toulouse III

☎ : 05.61.55.67.81 - fax : 05.61.55.62.58 - mobile : 06.12.57.19.49

mèl : dousset@irit.fr – site web : <http://atlas.irit.fr>

Membres :

Philippe BENHAMOU, ONERA, Châtillon

Henri BRIAND, Université de Nantes

Claude CHRISMENT, IRIT-SIG, Toulouse III

Michel DORBAN, U. de Louvain la Nve (Belgique)

Jacques DUCLOY, CNRS/INIST, Nancy

Carlos E. ESCOBAR TOLEDO, UNAM, (Mexique)

Péré ESCORSA-CASTELLS, Iale Tec, Barcelone (Espagne)

Luc GRIVEL, Université de Paris I

Olivier JOUVE, SPSS, Paris

Geneviève LACOMBE, INRA, Sophia-Antipolis

Dominique LADIRAY, Statistic Canada, (Canada)

Alain LEFEVRE, GFI, (Belgique)

Humbert LESCA, ESA, Grenoble

Christian LONGEVIALLE, DGA, Paris

Ramón MASPONS, UPC, Barcelone (Espagne)

Nicolas MOINET, Université de Poitiers

Julio MONTI, Nestlé, Lausanne (Suisse)

Josiane MOTHE, IRIT-IUFM, Toulouse

Hiroshi MOTODA, Université d'Osaka (Japon)

Fionn MURTAGH, Ulster University (Irlande)

Eduardo OROZCO, Biomundi-Idict (Cuba)

Clément PAOLI, Université de Marne la Vallée

Serge QUAZZOTTI, Centre de V. Tech, (Luxembourg)

Marisela RODRIGUEZ, CETEC Monterrey (Mexique)

Maryse SALLES-COLLETIS, Université de Toulouse I

Henry SAMIER, Université d'Angers

Yang TIANJUN, UST Beijing (Chine)

Alessandro ZANASI, TEMIS SA, Modena (Italie)

Organisation :

Anne-Marie PESQUÉ - DELTA VEILLE

☎ : 05.53.77.88.33 - mobile : 06.87.01.67.35

fax : 05.53.77.88.34 - mèl : delta.veille@wanadoo.fr

Mise en relation de deux bases de données bibliographiques, PASCAL et le SCIENCE CITATION INDEX pour l'élaboration d'indicateurs de la production scientifique Brésilienne

Joachim QUEYRAS (*), Leandro INNOCENTINI LOPES DE FARIA (**),

Hervé ROSTAING (*), Luc QUONIAM (***)

pro.queyras@laposte.net , leandro@nit.ufscar.br , rostaing@crrm.u-3mrs.fr , quoniam@univ-tln.fr

(*) Université Paul Cézane Aix-Marseille III, UMR CNRS 6171, Faculté des Sciences et Techniques Service 422, 13397 Marseille Cedex 20, France,

(**) NIT/Materiais, Universidade Federal de São Carlos, UFSCar, Rodovia Washington Luis, km 235, CEP 13565-905, São Carlos, SP, Brasil,

(***) Université du Sud, Service N.T.I.C, Bât Le Béal - 1er étage, Avenue de l'Université - BP20132 83957 La Garde CEDEX, France.

Mots clés :

Base de données, Classification, Production scientifique, Bibliométrie

Keywords :

Database, Classification, Scientific production, Bibliometrics

Palabras clave :

Base dato, Clasificacion, Produccion cientifica, Bibliometria

Résumé :

Avec le développement considérable des systèmes informatiques, la mise en relation de sources d'informations toujours plus importantes devient un des objectifs clés pour l'analyse informationnelle et le développement des systèmes d'information. Cet article présente la méthodologie de la mise en relation, grâce à l'utilisation de la classification par domaines scientifiques de l'ISI, de deux bases bibliographiques, la base PASCAL et le Science Citation Index Expanded (SCIE). L'analyse, par des méthodes bibliométriques, de la production scientifique brésilienne des deux bases au cours de la période allant de 1998 à 2002, a permis l'élaboration d'indicateurs à haute valeur ajoutée. Enfin, ce travail s'inscrit dans le cadre de la réalisation de l'annuaire statistique de la FAPESP¹ : « *Indicadores de Ciência, Tecnologia e Inovação em São Paulo – 2003* » et dont la publication est prévue pour la fin de l'année 2004.

¹ Fundação de Amparo à Pesquisa do Estado de São Paulo (Brésil)

1 Introduction

La mondialisation et le développement fulgurant des technologies de l'information et de la communication ont généré ces dernières années une augmentation constante de la masse d'information et une accélération des phénomènes d'ordre économiques et politiques. L'instabilité croissante qui en résulte a provoqué de profondes mutations dans l'organisation des entreprises qui ont de fait développé des systèmes de veille stratégique et scientifique pour améliorer leur compétitivité. Dans ce contexte, la connaissance scientifique et technologique, au même titre que la connaissance de l'environnement, des marchés ou des concurrents, s'est révélée primordiale pour rendre la prise de décision plus efficiente.

La production d'information à haute valeur ajoutée est donc primordiale et entraîne le développement de nouvelles méthodes d'extraction et d'analyse des données qui aboutissent, le plus souvent, à l'élaboration d'indicateurs. Ces derniers, par leur diversité, permettent de construire des représentations compréhensibles de manifestations toujours plus complexes et d'affiner, par la même, la perception et l'analyse des décideurs. Aussi, pour leur construction, l'utilisation commune de plusieurs sources d'information s'avère indispensable pour élargir le champ de leur description.

Aussi, avec le développement considérable des systèmes informatiques, la mise en relation de sources d'informations toujours plus importantes, des bases de données à plusieurs millions d'entrées par exemple, devient un des objectifs clés pour l'analyse informationnelle et le développement des systèmes d'information. Cet article présente la méthodologie de la mise en relation, par la construction d'une table de descripteurs communs, de deux bases bibliographiques importantes, la base *PASCAL*² et la *Science Citation Index Expanded*³ (*SCIE*).

Nous décrivons donc, dans un premier temps, la méthodologie que nous avons suivie pour la mise en relation de ces deux bases bibliographiques (1), puis, nous exposerons quelques résultats de l'analyse bibliométrique des articles scientifiques publiés par des chercheurs brésiliens dans chacune des deux bases (2) afin de souligner l'importance de leur utilisation commune.

2 METHODOLOGIE

2.1 Les bases de données

Les bases de données représentent des univers fermés. Il est donc possible de fabriquer des indicateurs à partir de plusieurs bases qui ne soient en aucunes mesures joignables. Pour travailler avec plusieurs bases de données, il est donc indispensable de dégager un langage commun. Si l'on regarde les bases de données de l'*ISI*⁴ et la base *PASCAL*, seuls les champs « contenants », c'est à dire qui décrit les références incluses dans les bases de données (journal de publication, affiliation, volume...), sont communs. Or, il est important de travailler sur des indicateurs de contenu (titre, mots clés, classification...) qui doivent également être rapprochables et comparables. La problématique de ce travail consiste, finalement, à fabriquer un vocabulaire commun qui permette de relier le contenu des deux bases.

Les critères de sélection des périodiques, indexés dans les bases de données utilisées pour l'élaboration d'indicateurs de la production scientifique d'un pays, influence fortement les résultats de l'analyse [ZITT, 2003]. La description de la production scientifique du Brésil, finalité de ce travail, réalisée à partir de l'analyse bibliométrique de deux corpus de références extraits respectivement de la base *PASCAL* et du *SCIE*, présente donc deux visions complémentaires de la science au Brésil dans l'analyse des résultats.

² Base de données produite par l'Institut National de l'Information Scientifique et Technique

³ Base de donnée produite par l'Institute for Scientific Information

⁴ Institute for Scientific Information (EUA)

PASCAL est une base de données multidisciplinaire, multi affiliation et multilingue qui intègre, depuis 1973, de la littérature mondiale en Science, Technologie et Médecine. Elle est produite par l'*INIST*⁵ et compte aujourd'hui 14,7 millions de références indexées à partir 6000 titres de journaux internationaux. C'est l'unique base de données multidisciplinaire de taille comparable au *SCIE* qui, traditionnellement, complète la vision « américaine » de la science exprimée par les bases de l'*ISI* par une vision plus « européenne ».

De même, le *SCIE* est une base de données multidisciplinaire et multi affiliation qui couvre, depuis 1945, les publications scientifiques mondiales, principalement dans les domaines des sciences exactes et des sciences biologiques. Elle est produite par l'*ISI* et indexe aux alentours de 6.068 périodiques scientifiques, ce qui représente plus de 26 millions d'articles scientifiques.

Les visions apportées respectivement par ces deux bases de données sont complémentaires, car elles offrent des corpus de références différents pour l'analyse de mêmes phénomènes. Aussi, en admettant que la concordance des indicateurs produits par des bases différentes agit comme validation de l'analyse et que la discordance alerte sur la nécessité de la réalisation d'études complémentaires [JAGODZINSKI-SIGOGNEAU, 1991], la production d'indicateurs à partir de ces deux bases est donc plus fiable.

2.2 La collecte et le reformatage

La collecte des données a été réalisée dans les deux bases de manière à recueillir l'ensemble de la production scientifique brésilienne de la période. 65066 références (31164 pour la période allant de 1998 à 2002) ont ainsi été extraites de l'ensemble de la base *PASCAL* contre 79.136 (période 1998 à 2002) pour le *SCIE*. Ont été pris en considération tous les articles dans lesquels au moins un auteur était affilié à une institution brésilienne au moment de la publication. La préparation des données, qui précède l'élaboration des indicateurs, a permis de normaliser l'ensemble des références (reformatage) et de les réorganiser dans de nouveaux champs de manière à optimiser leur analyse bibliométrique. Cette étape a été automatisée grâce à l'utilisation du programme *Infotrans*.

2.3 Quelle classification ?

La segmentation de la production scientifique par domaines de connaissance est fondamentale pour la production d'indicateurs. Dans le cas de l'étude générale de la production scientifique d'un pays, cette classification doit être suffisamment large pour permettre de créer des macro indicateurs sur la globalité des références scientifiques considérées. Elle doit également être intégrée dans chacune des références scientifiques qui vont servir de données initiales pour l'analyse. Plusieurs classifications ont été prises en compte pour la réalisation de ce travail :

- La *National Science Foundation* emploie, pour l'élaboration de son annuaire statistique, le « *Science & Engineering Indicators* » une classification qui segmente, en 9 grands domaines de connaissance, la production scientifique mondiale. La méthodologie employée pour élaborer cette classification, à partir des publications indexées dans les bases de l'*ISI*, n'étant pas disponible, nous n'avons pas pu l'employer pour la classification de nos corpus.
- Une autre classification, moins générale, a été développée récemment pour l'élaboration des *Essential Science Indicators (ESI)*⁶, produit de l'*ISI*. Elle reste très générale puisqu'elle sectionne en 22 domaines la connaissance scientifique.
- Le plan de classement développé par l'*INIST* pour le référencement des articles scientifiques est très complet puisqu'il associe à 8 grands domaines scientifiques une multitude de sous domaines. Mais les nombreuses modifications qu'il a subies rendent son utilisation impossible pour le classement de références scientifiques publiées au cours d'une longue période.

Cela étant, il existe des différences importantes qui doivent être soulignées pour permettre, par la suite, une meilleure interprétation des indicateurs. La première différence concerne le mode de classification des données à l'intérieur de chaque base. L'*INIST* procède à la classification de chacune des

⁵ Institut National de l'Information Scientifique et Technique

⁶ www.isinet.com/products/evaltools/esi/

références, indépendamment les unes des autres. A l'inverse, les références des bases de données de l'*ISI* sont classées par l'intermédiaire de leurs journaux d'origine, indexés dans la base, et auxquels ont été attribués des domaines de connaissance. Ainsi, il est possible de rencontrer dans une même référence de *PASCAL* plusieurs items du plan de classements de l'*INIST*, auxquels se réfère cet article, alors qu'aux références du *SCIE* ne sera attribuée, dans la majorité des cas, qu'une classification introduite à partir du journal de publication.

Pour établir la relation entre les deux bases, le plus simple est alors de considérer les mots-clés les plus génériques. Or, l'*ISI* décrit les journaux de publication, et non les articles, comme le fait la base *PASCAL*. La classification de ces journaux peut-être considérée dès lors comme le niveau de description le plus général des deux bases. C'est pourquoi nous avons retenu la classification de l'*ESI* comme classification pour les deux bases.

Enfin, la description de chaque article du corpus de références de la base *PASCAL* (65066 références) par la classification de l'*ESI* revient à construire la table de relation thématique entre le *SCIE* et la base *PASCAL*, y compris pour les journaux de publication qui ne sont pas pris en compte dans les bases de données de l'*ISI*.

2.4 La classification des références

Pour pouvoir comparer les indicateurs obtenus à partir des deux bases de données, il est nécessaire, pour les domaines de connaissance des références scientifiques, d'utiliser une classification commune.

Medecine	Mathématiques
Physique	Sciences de la Terre
Science de la Vie et de la Nature	Génétique et Biologie moléculaire
Chimie	Pharmacologie et Toxicologie
Biologie et Biochimie	Sciences de l'Univers
Ingénierie	Immunologie
Neurosciences et Comportement	Informatique
Sciences agricoles	Multidisciplinaire
Science des matériaux	Psychiatrie et Psychologie
Ecologie	Sciences Sociales
Microbiologie	Economie
Source : Essentials Science Indicators - ISI, [2004]	

Tableau 1 : Classification *ESI* des domaines de connaissance

La classification à 22 items de l'*ESI* (Tableau 1) semble la mieux adaptée à la réalisation de ce travail car il est possible, à partir des informations disponibles sur le site de l'*ISI*, de créer la liste de correspondances entre les journaux indexés dans le *SCIE*, et donc de leur ISSN, et la classification par grands domaines de connaissance. On peut ainsi intégrer, grâce à l'utilisation du programme *Infotrans*, cette classification à l'intérieur de chaque référence de la base de données.

Or, il existe de nombreux journaux qui sont indexés à la fois dans *PASCAL* et dans le *SCIE*. Il est donc possible, pour ces journaux tout au moins, d'intégrer, dans les références de *PASCAL*, la classification de l'*ESI* déjà incorporée aux références du *SCI*. Il suffit pour cela de substituer, dans chaque référence, l'ISSN par la classification correspondante, comme cela se rencontre déjà dans le *SCIE*. Pour illustrer ce propos, sur la période de publication allant de 1998 à 2002, 26.379 registres, sur 31.164, ont été classifiés ainsi, soit environ 85% du corpus de références de *PASCAL* que nous avons utilisé pour faire l'étude de la production scientifique du Brésil.

Classification PASCAL	Classification SCIE														
	Physique	Médecine	Chimie	Science de la Vie et de la Nature	Ingenierie	Science des Matériaux	Neurosciences et Sciences du Comportement	Pharmacologie et Toxicologie	Ecologie	Biologie et Biochimie	Microbiologie	Sciences Agricoles	Sciences de la Terre	Informatique	Mathématique
ETAT CONDENSE : STRUCTURE ELECTRONIQUE, PROPRIETES ELECTRIQUES, MAGNETIQUES ET OPTIQUES.	0,30	0,02			0,04										
PHARMACOLOGIE		0,09	0,03	0,02		0,05	0,24		0,01	0,02	0,01				
PATHOLOGIE		0,14		0,02							0,08				
ETAT CONDENSE : STRUCTURE, PROPRIETES MECANIQUES ET THERMIQUES	ETA1		0,07		0,07										
PHYSIQUE	0,15				0,06									0,01	0,02
DOMAINES INTERDISCIPLINAIRES : SCIENCE DES MATERIAUX, RHEOLOGIE	0,10				0,10	0,03								0,02	0,01
AGRONOMIE					0,10			0,13		0,08					
CHIMIE	0,02		0,20	0,01	0,05	0,04									
ECOLOGIE					0,18			0,19		0,02		0,02			
INFORMATIQUE					0,15									0,31	0,01
MATHEMATIQUES					0,11									0,06	0,36
MICROBIOLOGIE		0,02		0,03					0,02	0,22					
PHYSICOCHIMIE DES POLYMERES	0,02		0,18		0,04										
BIOTECHNOLOGIE			0,02	0,02				0,04	0,28	0,04	0,01				
CHIRURGIE		0,12													
PHYSIQUE : SCIENCE DES MATERIAUX	0,05	0,02			0,17										

Source : Essentials Science Indicators - ISI, [2004]

Tableau 2 : Matrice partielle de corrélation entre la classification de PASCAL et du SCIE

A partir de ces 26.379 références, nous avons construit, par l'utilisation du programme *VantagePoint*, la matrice de correspondance entre les deux classifications. Nous nous sommes limités aux six premiers caractères de la classification *INIST* pour réduire le nombre de formes à considérer. L'application de l'indice de Jacquard à l'ensemble des intersections de cette matrice a permis de déterminer la corrélation entre chaque item de chaque classification. Ainsi, pour chaque terme de la classification *INIST*, le meilleur indice de Jacquard a permis de déterminer l'item de la classification *ESI* à associer. La matrice de corrélation entre les deux classifications est présentée partiellement dans le Tableau 2.

Enfin, et sur l'ensemble du corpus de références extraites de *PASCAL*, la classification *ESI* a été intégrée de nouveau, mais, cette fois-ci, à partir directement des items du plan de classement de l'*INIST*. Seules 60 références, sur les 31.164 de notre corpus n'ont pas été classées. Elles correspondent à des codes de classement qui n'ont jamais été en corrélation avec la classification *ESI* et ne représentent qu'une infime partie qui n'entrera pas en considération pour l'élaboration des indicateurs.

2.5 Production des indicateurs

De très nombreux indicateurs ont été élaborés à partir des données extraites de la base *PASCAL* et du SCIE. Seules quelques figures et tableaux vous seront présentés dans ce travail, ils illustrent en effet l'importance de l'utilisation de plusieurs bases pour la réalisation d'une même étude.

Tous ces indicateurs ont été produits soit directement à partir du programme *VantagePoint* que nous avons utilisé pour construire les matrices, soit par le biais du logiciel Excel de la Suite Microsoft Office 2003, pour la réalisation des représentations graphiques.

3 RESULTATS

L'objectif de cette présentation n'est pas de faire la description de la production scientifique brésilienne. Nous exposerons, d'une part, les résultats de l'intégration de la classification de l'*ESI* dans les références de la base *PASCAL* et, d'autre part, quelques indicateurs qui soulignent l'importance de l'utilisation d'une même classification dans deux bases de données différentes.

3.1 Intégration de la classification ESI

Le tableau 3 présente la matrice de relation entre la classification *ESI* intégrée directement à partir des journaux de publication et intégrée par l'intermédiaire du plan de classement de l'*INIST*. Seules les références intégrant les deux classifications ont été prises en considération dans cette matrice. A un item de la classification intégrée par l'intermédiaire des journaux de publication sont associés plusieurs domaines scientifiques issus de l'intégration de la classification à partir des plans de classement de l'*INIST*. Deux éléments fondamentaux permettent d'expliquer cette différence. D'une part, chaque référence de la base *PASCAL* est classifiée indépendamment. Par conséquent, à chaque référence d'un même périodique ne sont pas forcément associés les mêmes items de la classification. D'autre part, l'*INIST* attribue aux références de la base *PASCAL* un ou plusieurs items de son plan de classement, ce qui explique l'augmentation du nombre de références associées à un même domaine scientifique, mais aussi la possibilité, pour un article référencé dans la base *PASCAL*, d'être répertorié par plusieurs items de la classification par domaines scientifique de l'*ESI*.

	# Records	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	# Records	1281	8183	4311	4201	4178	3686	3284	3227	2644	2182	1987	1881	1854	1672	1383	1209	677	842
CLASSIFICACAO	Records	PHYSICS	CLINICAL MEDICINE	PHARMACOLOGY & TOXICOLOGY	CHEMISTRY	ENVIRONMENTAL ECOLOGY	MATERIALS SCIENCE	BIOLOGY & BIOCHEMISTRY	ENGINEERING	PLANT & ANIMAL SCIENCE	COMPUTER SCIENCE	NEUROSCIENCE & BEHAVIOR	MECHANICS	MINERALOGY	MATHEMATICS	MOLECULAR BIOLOGY & GENETICS	AGRICULTURAL SCIENCES	SPACE SCIENCE	SOCIAL SCIENCES, GENERAL
1	18115	PHYSICS	9220	44	7	402	47	1124	74	582	56	58	32	84	89	70	6	284	8
2	8191	CLINICAL MEDICINE	8	8415	1360	2	51	3	715	2	53	22	362	275	462	1			306
3	4782	CHEMISTRY	8561	15	2367	2340	122	544	225	1621	44	119	1	41	7	17	13	128	
4	2879	PLANT & ANIMAL SCIENCE	23	194	136	77	1570	80	278	6	1889	64	12	57	184	2	162	64	120
5	3807	ENGINEERING	922	56	23	523	173	315	51	1776	21	1018	21	182	1	584	21	47	14
6	2292	MATERIALS SCIENCE	740	22	23	322	17	1450	10	175	21	9	2	1		4	2		
7	1882	NEUROSCIENCE & BEHAVIOR	132	331					101	21	4	1176			1	4	9	2	
8	1844	PHARMACOLOGY & TOXICOLOGY	66	1562	5	7		91		16	4	148		14		62			
9	1828	ENVIRONMENTAL ECOLOGY	3	13	84	19	1559	6	147	33	120	2	1	305	20	3	17	56	12
10	1775	BIOLOGY & BIOCHEMISTRY	1	148	96	3	82	14	1182	2	97	1	35	76		117	35		17
11	1611	MICROBIOLOGY		392	129		70		138		166	1	2	2	353	72	13		41
12	1326	AGRICULTURAL SCIENCES		31	84	8	290	32	116	11	97	2	7	60	8	1	12	805	23
13	1264	GEOSCIENCES	1	7	2	2	178	78	4	138	3	2						37	9
14	1077	COMPUTER SCIENCE	101	7		6	8	21	5	227	5	822	4	18	1	173	13	3	1
15	722	MATHEMATICS	100	2	1		3		58	4	45	2	2						1
16	724	MINERALOGY		359	56				11	11	13				325	592			
17	712	SPACE SCIENCE		24		1	2			2			42	25					685
18	617	MOLECULAR BIOLOGY & GENETICS		53	61				57	2	13					286			10
19	428	SOCIAL SCIENCES, GENERAL		249	22		5	1	11	45	28		11	6		6		1	201
20	381	PSYCHIATRY/PSYCHOLOGY		43	37				3	3		108							1
21	73	MULTIDISCIPLINARY	12	2	3	1	13	2	4				2	16	6			9	1
22	36	ECONOMICS & BUSINESS							24		8	2				6			1

Tableau 3 : Matrice de relation entre la classification ESI intégrée directement à partir des journaux de publication et intégrée par l'intermédiaire du plan de classement de l'INIST

3.2 Vers un système d'information transversal

La construction de la table de relation entre le SCIE et la base PASCAL permet de relier deux sources d'informations académiques intéressantes et complémentaires. Cette base de données ayant déjà été reliée à la Classification Internationale des Brevets [FAUCOMPRE, QUONIAM 1997], et à la terminologie des codes de classification du registre d'entreprises Kompass, elle se place au centre d'un véritable système d'information transversal et performant auquel devrait certainement pouvoir se greffer d'autres sources d'information, au niveau national mais également international.

3.3 Indicateurs de la production scientifique brésilienne

N° de publicações/ Área do conhecim.	N° de publicações por região geográfica					
	Sudeste	Sul	Nordeste	Centro-oeste	Norte	97
1660	Total	957	186	192	41	10
Fisica	USP	UFRGS	UFPE	UNB	UFAM	
	UNICAMP	UFSC	UFC	UFG	UFRR	
	UNESP	UFPR	UFPB	PUC BRASIL.		
	Total	532	97	42	33	8
Medicina	USP	UFRGS	UFBA	UNB	INPA	
	UNIFESP	UFPR	UFC	UFG	UFAC	
	UFRJ	UFSC	UFPE	UFMT		
	Total	281	106	24	54	21
Ecologia e meio ambiente	USP	UFRGS	UFPE	EMBRAPA	INPA	
	UNICAMP	UFSC	UFPB	UNB		
	UFV	UFMS	UFC	UFMG		
	Total	365	76	127	11	9
Quimico	USP	UFRGS	UFPE	EMBRAPA	UFRR	
	UFRJ	UFSC	UFRN	UNB	UFAM	
	UNICAMP	UFPR	UFPB			
	Total	331	103	48	18	6
Farmacologia e Toxicologia	USP	UFRGS	UFC	UFG	UNIPAR	
	UNIFESP	UFSC	UFPE	UNB		
	UFRJ	UFPR	UECE			
	Total	268	54	39	10	4
Ciência dos materiais	USP	UFRGS	UFPE	UNB	INPA	
	UFSCAR	UFPR	UFC	EMBRAPA		
	UNICAMP	UFSC	UFPB	UFG		

Tableau 4 : Nombre de publications et principales Institutions brésiliennes par région et domaine de connaissance (extrait)

Le tableau 4 présente un indicateur qui a été élaboré à partir de la base PASCAL pour l'année 2002. Il relève, pour chaque région brésilienne, le nombre de publications et les trois principales institutions se rapportant à chaque domaine de connaissance. Cet indicateur permet de souligner l'importance du reformatage des données puisqu'à partir d'un corpus de 31.164 références, il est possible de produire un méso indicateur permettant de considérer l'importance des propres institutions dans la production scientifique brésilienne par région et par domaine scientifique.

Le tableau 5 mesure le nombre de publications par domaine scientifique dans les corpus extraits de la base PASCAL et du SCIE.

Dans chacune des deux bases de données, l'analyse du nombre de publications par domaine scientifique offre des résultats très différents. S'il est évident d'affirmer, analysant les résultats obtenus à partir des deux bases, que la majorité des publications brésiliennes indexées sont répertoriées dans les domaines de la physique, de la chimie et de la médecine, il est plus difficile de déterminer laquelle de ces trois disciplines englobe le plus de publications.

Finalement, cet indicateur est important pour démontrer qu'il n'existe pas qu'un seul résultat, mais plusieurs qui doivent être regroupés pour permettre de décrire au mieux un phénomène. Dans le cas des bases de données PASCAL et le SCIE, par exemple, la couverture scientifique de chacune d'elle varie puisqu'elles indexent des journaux pour partie différents. Cela engendre la production de résultats variables, qui ne s'opposent pas, mais qui, au contraire, sont complémentaires. Enfin, Pour comprendre plus en détail ces indicateurs, il serait important de mener, parallèlement, d'autres études, à partir de sources d'information encore différentes, pour pouvoir compléter l'interprétation de ces résultats. D'autant plus que la rigueur méthodologique qui a accompagné cette analyse ne permet pas de supposer l'existence de biais lors des différents traitements. La multiplication des approches pour l'étude d'un même phénomène, alors observé selon différents points de vue, permet, finalement, d'en améliorer son analyse.

Science Citation Index Expanded - Web of Science*		Pascal via Dialog On Disc**	
ÁREAS DE CONHECIMENTO			
Medicina	2 848	Física	642
Física	1 052	Medicina	444
Química	970	Química	247
Ciências dos animais e plantas	573	Farmacologia e toxicologia	230
Biologia e bioquímica	539	Ciência dos materiais	207
Engenharia	388	Ecologia e meio ambiente	170
Neurociência e comportamento	378	Biologia e bioquímica	159
Ciência dos Materiais	282	Geociências	151
Biologia molecular e genética	212	Neurociência e comportamento	149
Ciências agrárias	191	Microbiologia	109

*Fonte : SCIE - ISI via Web of Science, [2004]

**Fonte : Pascal - INIST via Dialog On Disc, [2004]

Tableau 5 : Nombre de publications par domaine scientifique pour l'année 2002. Base SCIE et PASCAL

4 CONCLUSION

Dans cet article, nous proposons une méthodologie pour l'élaboration de la table de relation entre le SCIE et la base PASCAL qui se place, finalement, au centre d'un véritable système d'information performant et transversal allant de la recherche fondamentale, aux secteurs industriels et appliqués. De plus, en décrivant le système qui permet de mettre en relation des bases de données de plusieurs millions de registres, cet article ouvre la voie à l'élargissement des systèmes d'information pour l'élaboration d'indicateurs toujours plus complets.

Ce travail s'inscrit dans le cadre de l'analyse de la production scientifique brésilienne au cours de la période allant de 1998 à 2002 dont les résultats seront publiés dans l'annuaire statistique de la FAPESP « *Indicadores de Ciência, Tecnologia e Inovação em São Paulo – 2003* » et dont la publication est prévue pour la fin de l'année 2004.

Enfin, nous remercions chaleureusement l'INIST et DIALOG BRAZIL sans qui ce travail n'aurait jamais pu voir le jour.

5 Bibliographie

- CNPq. Diretório dos Grupos de Pesquisa. Disponible à l'adresse <<http://lattes.cnpq.br/diretorio/>>. Consultado em 10/01/2004
- COURTIAL, J. P. Introduction à la scientométrie: de la bibliométrie à la veille technologique. Paris: Anthropos, 1990.
- EC. European Comission. Directorate-General for Research. Third European Report on Science & Technology Indicators-2003. Disponible à l'adresse <ftp://ftp.cordis.lu/pub/indicators/docs/3rd_report.pdf>
- FAPESP. Indicadores de C & T e inovação em São Paulo – 2001. organização: Francisco Romeu Landi. São Paulo: Fapesp, 2002, p.488
- FARIA, L. I. L.; QUONIAM, L.; MUGNAINI, R. Elementos de comparação das bases de dados Pascal e Scisearch. ISDM, n.5, 13p, Dez/2002. Disponible à l'adresse http://isd.m.univ-tn.fr/articles/num_archives.htm
- FAUCOMPRES, P.; QUONIAM, L.; DOU, H. The function-application relation through a link between classification and indexing. *World Patent Information*, England, Vol. 19, No. 3, pp. 167-174, 1997.
- INIST. Pascal – Description. Disponible à l'adresse <<http://www.inist.fr/en/PRODUITS/pascal.php>> Consultado em 15/01/2004.
- JAGODZINSKI-SIGOGNEAU, M. et al. Scientific innovation in bibliographical databases: a comparative-study of the Science Citation Index and the Pascal Database. *Scientometrics*, v. 22, n. 1, p. 65-82, 1991.
- LUWELL, M. Is the Science Citation Index US-biased? In: CONFERENCE OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS, 7, 1999, Colima, MX. Proceedings... Colima, MX: International Society for Scientometrics and Informetrics. 1999. 303-312.

- NSB. National Science Board. Science and Engineering Indicators. Arlington, VA: National Science Foundation, 2002
- OST. Observatoire des Sciences et des Techniques, Les chiffres clés de la science et de la technologie, Economica, 2003
- ROSTAING, H. La bibliométrie et ses techniques. Collection "Outils et méthodes", co-édition sciences de la société et CRRM - Centre de Recherche Rétrospective de Marseille. Marseille. 1996
- SCIELO. Disponível em <<http://www.scielo.br>>. Consultado em 25/01/2003.
- SPINAK, E. Dicionário enciclopédico de bibliometria, cienciometria e informetria. Caracas: UNESCO CII/II, 1996.
- SPINAK, E. Indicadores cienciométricos. Ciência da Informação, Brasília, DF, v. 27, n. 2, p. 141-148, 1998.
- TESTA, J. A base de dados ISI e seu processo de seleção de revistas. Ciência da Informação, Brasília, DF, v. 27, n. 2, p. 233-235, 1998.
- VELHO, L. Cuidado com os rankings científicos. Disponible à l'adresse <<http://www.prometeu.com.br>>. Dernier accès le 20 août 2001.
- VIOTTI, E.B. e MACEDO, M.M. (org.) - Indicadores de C & T e inovação no Brasil. Campinas, SP: Editora UNICAMP, 2003, p.614
- ZHU, D.; PORTER, A. L. et al. A process for mining science & technology documents databases illustrated for the case of knowledge discovery and data mining. Ciência da Informação, Brasília, DF, v. 28, n.1, jan. 1999.
- ZITT, M.; RAMANANA-RAHARY, S.; BASSECOULARD, E. - Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation. Scientometrics. Netherlands, v. 56, n. 2, pp 259-282, 2003.