

Etude bibliométrique pour l'évaluation des programmes de recherche nationaux - Difficulté de mise en œuvre et d'exploitation dans le cas de la recherche scientifique algérienne

HERVÉ ROSTAING, VALÉRIE LÉVEILLÉ

rostaing@crrm.u-3mrs.fr, leveille@crrm.u-3mrs.fr

CRRM, Centre scientifique de Saint-Jérôme, Université Aix-Marseille III, 13397 Marseille cedex 20

tel : 33~(0)4 91 28 87 46 / fax : 33~(0)4 91 28 87 12 / web : <http://crrm.u-3mrs.fr>

Mots-clés : bibliométrie, évaluation de la recherche, programmes de la recherche nationaux, nomenclature, normalisation des données, agrégation des données, plan de classement

Keywords: bibliometrics, research evaluation, national research programs, nomenclature, data standardization, data classification, classification scheme

Palabras Clave: bibliometría, evaluación de la investigación, programas de investigación nacionales, nomenclatura, normalización de los datos, clasificación de los datos, tablas de clasificación

Résumé : Les études bibliométriques ont pour but d'aider à la compréhension de grands volumes de données scientifiques et techniques. Cet article expose les réflexions et les traitements à mettre en œuvre pour faciliter l'appropriation des résultats des études bibliométriques. Les techniques bibliométriques employées pour l'évaluation de recherche au niveau national sont particulièrement sensibles à la phase de recodification des données brutes de façon à favoriser cette appropriation. Pour ce faire, le choix d'une source de données adaptée, le besoin d'une stratégie d'agrégation et la nécessité d'adaptation des traitements aux référentiels de lecture des destinataires sont les garants de résultats pertinents. Ces réflexions ont été menées lors de l'étude de l'évaluation de la recherche scientifique en Algérie pour la période 1990-1999. Le choix de la base de données et de la stratégie d'agrégation des données est présenté. L'article expose plus précisément la nomenclature choisie pour qualifier les domaines scientifiques et techniques des travaux algériens. Les références ont été requalifiées sur la base d'une table de correspondance entre le plan quinquennal de la programmation de la recherche algérienne et le plan de classement de la base de données traitées. Cette recodification des données brutes a permis de présenter les résultats bibliométriques selon les programmes de recherche prioritaires définis dans le plan national. La représentation de la recherche algérienne n'en est que plus facilement acceptée, compréhensible et utile aux décideurs algériens.

Les techniques bibliométriques pour la veille et pour les évaluations nationales

Comment livrer les résultats d'une étude bibliométrique de la manière la plus adaptée aux attentes des destinataires ? Comment s'assurer que les informations obtenues par les traitements bibliométriques pourront être lues et interprétées ? Comment reformuler au plus juste les données brutes et les résultats statistiques pour favoriser leur acceptation et leur exploitation ? Bien souvent, lors de la mise en œuvre de traitements bibliométriques dans le cadre d'une action de veille la question ne se pose pas tout à fait en ces termes. L'analyse et l'interprétation des résultats bibliométriques est le résultat d'une collaboration entre le ou les spécialistes en information et le ou les experts du domaine scientifique ou technique étudié. Pour que l'action de veille porte ses fruits, ces experts sont choisis en fonction de leur degré de connaissances et/ou de compétences pour le sujet étudié. Ils sauront parfaitement comprendre les informations qui leur seront livrées et pourront ainsi les interpréter. La terminologie utilisée lors des traitements pour décrire les aspects scientifiques et techniques n'aura bien souvent aucun mystère pour eux. Même si les termes ne sont pas exactement ceux employés dans leur organisation, ils sauront rapidement les traduire par leur connaissance du vocabulaire utilisé dans la communauté internationale dont ils relèvent. Le groupe de veille aura tout loisir par la suite de n'exploiter que les faits marquants de cette analyse et de les reformuler pour construire un discours adapté aux destinataires de l'action de veille.

Dans le cadre d'une action de veille, les deux étapes de décodage puis recodage de l'information sont intrinsèque au processus et ne laissent que très rarement de trace lors de la restitution finale de l'étude. L'étude bibliométrique exposée dans cet article, ne répond pas du tout aux mêmes objectifs. Les techniques bibliométriques employées ici rentrent dans un projet d'analyse bien plus vaste. Elles vont

être utilisées pour l'évaluation de la recherche d'un pays. Comment mesurer le potentiel de recherche d'un pays et son dynamisme ? Comment montrer ses points forts et points faibles ? Comment analyser sa structuration, ses réseaux de collaborations ? Comment tenter de rendre plus efficiente la politique de la recherche publique ? Autant de questions que bon nombre de nations se posent. Les études bibliométriques sont souvent employées pour aider à répondre à ces interrogations. Il devient difficile d'imaginer, comme cela est fait dans une action de veille, de faire appel à des experts pour chaque domaine scientifique et technique couvert par la recherche nationale. Il faut donc s'abstenir du soutien de ces "interprètes" nous permettant de décoder les données brutes. Par contre, les destinataires sont tout autant exigeants que les décideurs des entreprises. Il faut leur livrer non seulement des résultats fiables mais aussi compréhensibles. Les décideurs de la recherche publique ne sont pas plus disposés que les décideurs des entreprises à perdre du temps pour l'assimilation de résultats trop "indigestes". Le besoin de recoder l'information pour le destinataire reste toujours autant essentiel.

Les nomenclatures dans les techniques bibliométriques appliquées aux évaluations nationales

Lors de l'application de techniques bibliométriques pour l'évaluation de la recherche nationale plusieurs contraintes se font sentir :

- Une source de données adaptées : très peu de bases de données peuvent faire l'objet d'une analyse bibliométrique à caractère nationale. Seules deux bases de données ont à la fois une couverture internationale et multidisciplinaire (en sciences) : la base américaine Science Citation Index (SCI) de l'ISI¹, et la base française Pascal de l'INIST².
- Le besoin d'agrégation des données brutes : les besoins d'analyse des décideurs de la politique de la recherche nationale ne nécessitent pas que les éléments descriptifs de cette recherche soient d'une très grande finesse. Or, les données brutes employées en bibliométrie, les références bibliographiques, font appel à une granularité de ces descripteurs bien trop fine. Les bases de données étant conçues à des fins de recherche documentaire et non pas d'analyses statistiques nationales, les données brutes devront subir des traitements d'agrégation avant de pouvoir offrir des enseignements utiles à une investigation nationale.
- Adaptation des résultats des traitements bibliométriques aux référentiels de lecture des destinataires de l'étude : les agrégations et les méthodes bibliométriques appliquées doivent être pensées selon les capacités de compréhension des destinataires et selon le référentiel de lecture qu'ils ont l'habitude d'employer ou de rencontrer. Il paraît inutile de livrer des cartes d'analyse de données multidimensionnelles à des personnes non initiés comme il paraît vain d'offrir comme grille de lecture des résultats basés sur des entités ou des concepts qu'elles n'ont pas l'habitude et le besoin de connaître.

En conséquence, les études bibliométriques mises en œuvre pour des évaluations nationales emploient des nomenclatures pour définir les règles d'agrégation des données brutes. La plupart des travaux déjà réalisés font références soit à la nomenclature proposée par l'ISI soit à la version revisitée par le CHI³. Ces deux nomenclatures sont construites à partir de grandes catégories thématiques attribuées par l'ISI à chaque référence introduite dans ses bases de données. Les références bibliographiques contenant les données utiles à l'établissement de ces nomenclatures, le travail s'en trouve fortement allégé. Par contre, les bibliométriciens se voient contraint d'exploiter une seule et même source pour toutes leurs analyses de la recherche scientifique : la base SCI (ou sa version améliorée par le CHI). Ainsi, toutes les analyse emploient les mêmes règles d'agrégation des données (la vision qu'à l'ISI de la science) au dépend des attentes spécifiques de chaque pays. Néanmoins, cette pratique monopolistique de l'ISI offre au moins l'avantage de pouvoir comparer les résultats des études menées par les différentes équipes internationales.

¹ Institute for Science Information [<http://www.isinet.com>]

² Institut pour l'Information Scientifique et Technique [<http://www.inist.fr>]

³ Computer Horizon Inc [<http://www.chiresearch.com>]

L'évaluation de la recherche scientifique Algérienne

Dans le cadre de l'étude bibliométrique menée pour l'évaluation de la recherche scientifique en Algérie, nous n'avons pas opté pour la pratique traditionnelle de l'utilisation de la base SCI de l'ISI.

De part le caractère historique et linguistique du rapprochement entre la France et l'Algérie, nous avons estimé qu'il était fort probable que l'activité scientifique algérienne était mieux prise en compte par l'institut français. Cette hypothèse n'a pas encore été vérifiée mais une étude comparative sera prochainement menée entre les résultats obtenus à partir de Pascal et du SCI.

Un second argument est venu plaider en faveur de l'exploitation de la base Pascal. Les éléments descriptifs du contenu scientifique des travaux référencés dans Pascal sont bien plus précis que ceux offerts par le SCI. La base Pascal dispose à la fois d'un plan de classement documentaire et de mots-clés comme éléments descriptifs contrôlés pour chaque article alors que l'ISI ne dispose que d'une catégorisation (Journal Subject Category) attribuée selon le thème du journal où l'article a été publié et aucun élément descriptif contrôlé de l'article en lui-même (les Identifiers Keywords n'ont été introduits qu'en 1991).

Le choix de la base Pascal établi comme source des données, la stratégie d'analyse d'agrégation des données s'est posée. Il a été établi que la recherche algérienne pourrait être appréciée selon cinq dimensions :

- Les acteurs algériens,
- Les collaborations internationales,
- Les domaines scientifiques et techniques,
- La stratégie de communication des travaux,
- Le temps.

La dimension *Temps* ne nécessite pas d'agrégation particulière puisque l'année de publication présente dans les références est la donnée à exploiter.

La dimension *Acteurs algériens* s'est exprimée selon trois niveaux de granularité : les auteurs algériens, les instituts algériens, les villes algériennes. L'obtention de ces informations n'a pas mis en œuvre de règles d'agrégation mais plutôt des traitements d'extraction et de normalisation des données à partir des champs bibliographiques Auteur et Affiliation.

La dimension *Stratégie de communication des travaux* peut s'exprimer à partir des données telles que le type du document, les journaux, les colloques et la langue de rédaction des travaux. Là encore, seuls des traitements de normalisation des données brutes ont été mis en œuvre.

La dimension *Collaboration internationale* a, par contre, fait appel à une nomenclature. Cette dimension a été exprimée selon trois niveaux de granularité : les instituts étrangers, les pays et un regroupement de pays selon leur proximité géographique. Les deux premiers niveaux de granularité nécessitaient uniquement des traitements de normalisation alors que le dernier a été créé à partir de la nomenclature exposée en Annexe 1. Il est à noter que nous n'avons pas utilisé les regroupements habituellement appliqués dans les évaluations de la recherche nationale (voir par exemple les nomenclatures de l'OST dans Barré R, 1995) car le point de vue de l'Algérie ne peut être le même que celui d'un pays du G7. Ainsi, nous n'avons pas regroupé la France à l'Europe de l'Ouest. Nous avons conservé une région Europe de l'Est. Nous avons voulu mettre l'accent sur les pays voisins de l'Algérie en créant une région Pays du Maghreb et en les détachant de la région Africaine.

Enfin la dimension *Domaines scientifiques et techniques* est celle que nous traiterons tout particulièrement dans cet article. Cette dimension impose l'usage d'une nomenclature adaptée. Les données brutes des bases de données n'offrent pas d'éléments décrivant les aspects scientifiques et techniques directement exploitables pour une évaluation nationale. Nous allons exposer dans les parties suivantes la nomenclature établie lors de l'étude de la recherche scientifique algérienne et les procédures appliquées au corpus de références pour le recoder selon ce nouveau référentiel.

Les données exploitées lors de l'étude

L'étude concerne uniquement les travaux algériens publiés pendant les dix dernières années (90-99). Cette période est très intéressante à évaluer car la politique de la recherche s'est structurée et affirmée vers une plus grande indépendance du modèle français résiduel. Il serait bon d'évaluer l'impact de cette politique sur les résultats de la recherche algérienne. Parallèlement, une situation de crise politique critique pour le pays est apparue lors de cette période. L'impact de cette crise sur l'activité scientifique algérienne serait source d'enseignements.

Les données exploitées pour cette étude bibliométrique ont été extraites de deux versions CD-ROM de la base Pascal puis complétées pour l'année 99 par la consultation d'une troisième version accessible par Internet :

- Les CD-ROM INIST-Jouve de 90 à 96 (version Française de Pascal),
- Les CD-ROM Dialog de 97 à 99 (version Anglaise de Pascal),
- La version web de Sylverplatter pour mettre à jour les références d'articles publiés en 99 mais seulement introduites dans la base à partir de 2000 (version Anglaise de Pascal).

A partir de ces sources, tous les articles comportant une affiliation algérienne (CS=ALGERIA ou AD=DZA) ont été sélectionnés pour constituer le corpus à étudier. Ce jeu de données a ensuite été normalisé (format unique, données dans la même langue) puis dédoublonné de façon à obtenir un corpus le plus fiable possible. Le corpus final obtenu est de 2545 références bibliographiques pour la période 90-99.

Ce jeu de données a ensuite subi plusieurs traitements d'extraction de données et de standardisation de façon à isoler ou créer de nouvelles données bibliométriques telles que : les noms des journaux et des colloques, le nom des pays, le regroupement des pays en région (Annexe 1), le nom des villes algériennes, le nom des instituts algériens et le nom des instituts étrangers. Tous ces traitements ont pu être automatisés grâce au logiciel Infotrans⁴.

Les programmes de la recherche algérienne

La construction des données bibliométriques représentant la dimension *Domaines scientifiques et techniques* a été la phase la plus longue dans cette étude.

Une première expérimentation élaborée à partir de l'expérience de l'application des études bibliométriques en veille technologique s'est avérée très rapidement infructueuse. Sachant que les mots-clés Pascal sont très peu pratiques pour une analyse bibliométrique (grande diversité, mélange entre générique et spécifique, ajout de descripteur non scientifique tel que des noms de pays ou zones géographiques...), nous avons décidé d'exploiter le plan de classement Pascal pour décrire le contenu scientifique et technique de chaque référence. Une rapide consultation⁵ de ce plan de classement montre que le premier niveau de la hiérarchie ne permet que de catégoriser les sciences en 2 macro-catégories (001 = Sciences Exactes et Technologies et 002 = Sciences biologiques et médicales). Le second niveau de la hiérarchie ne nous offre guère plus de satisfaction (exemple des sous-catégories de 001 : 001A = Sciences et techniques communes, 001B = Physique, 001C = Chimie, 001D = Sciences appliquées et 001E = Terre, océan et espace). Par contre, le troisième niveau du plan de classement devient trop précis. 97 catégories de ce 3^{ème} niveau ont été identifiées dans le corpus des travaux algériens (voir Annexe 3), avec des significations aussi variées que *Sciences de la terre* (code 001E01) et *Etat condensé: structure, propriétés mécaniques et thermiques* (code 001B60).

Ce troisième niveau de la hiérarchie n'étant pas assez agrégé pour certains aspects scientifiques, nous avons imaginé d'appliquer des regroupements de codes de façon à créer la nomenclature utilisée dans la plupart des études bibliométriques (nomenclature du CHI). Sur la base de travaux antérieurs (Barré R, 1991 et Narvaez N et al, 1999), une table de correspondances entre les codes Pascal et la nomenclature du CHI a été établie (voir Annexe 4). Deux catégories ont été rajoutées à la

⁴ Logiciel de reformatage développé par la société allemande IuK GmbH [<http://www.iukrieth.de/>]

⁵ Le plan de classification Pascal est consultable sur le site de l'INIST [<http://www.inist.fr/pascal/intropcpas.htm>]

nomenclature traditionnelle car il semblait impossible d'associer deux codes Pascal aux catégories existantes (*Santé publique. Hygiène-médecine du travail* et *Ethique*). Comme une seule de ces deux nouvelles catégories était présente dans les travaux algériens, la recherche algérienne se voit ainsi catégorisée selon 9 domaines. La répartition des publications algériennes sur la période 90-99 selon ces domaines est représentée par la Figure 1.

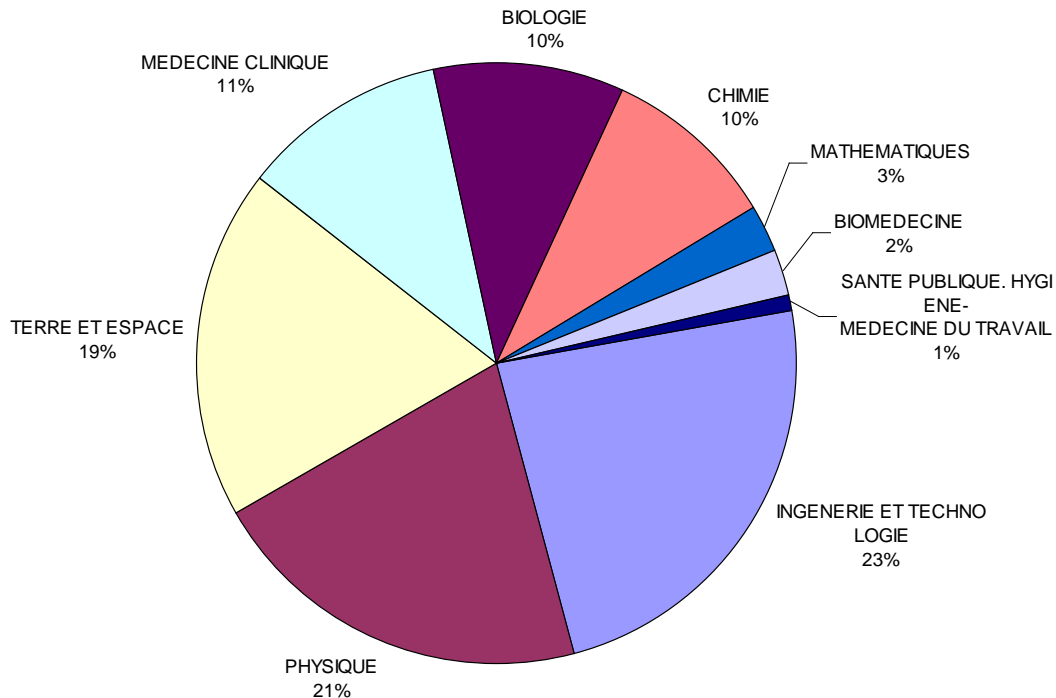


Figure 1 : répartition des publications algériennes par domaines selon la nomenclature du CHI

A la lecture de ce résultat, notre interlocuteur algérien nous a avoué qu'il ne retrouvait absolument pas les axes prioritaires de la recherche algérienne et qu'il lui était très difficile de le rapprocher des programmes de recherche planifiés par le gouvernement algérien. Dans le souci de privilégier l'appropriation des résultats bibliométriques par les destinataires algériens et dans le but de soumettre des résultats utiles à la politique de la recherche algérienne, nous avons alors utilisé une nomenclature propre à l'Algérie. Pour que cette nomenclature soit un référentiel commun à tous les acteurs impliqués dans la recherche algérienne, notre choix s'est porté sur le plan de programmation de la recherche établi par le gouvernement algérien.

Sur la base du texte de loi du dernier plan quinquennal de la programmation de la recherche (Loi n°98-11, 1998) et d'un rapport intermédiaire (Programmes nationaux de recherche, 1995), une table de correspondance entre le plan de classement Pascal et le plan quinquennal de programmation de la recherche a été élaborée. Ce plan fait apparaître 30 axes prioritaires dénommés *Programmes de recherche nationaux*. Chacun de ces programmes est subdivisé en *Axes*, chaque axe en *Domaines* et chaque domaine en *Sujets*. Parmi les 30 programmes de recherche, seulement 16 ont été retenus dans le cadre de notre étude (voir Annexe 2). Les autres programmes de recherche ne relevant pas de domaines scientifiques ou techniques aucun article les concernant n'est référencé dans la base Pascal. Pour obtenir une lecture plus fine de ces programmes de recherche, nous avons aussi recodé les articles du corpus en introduisant non seulement une codification de ces programmes (voir deuxième colonne du tableau en Annexe 2) mais aussi les axes qualifiant ces programmes.

L'élaboration de la table de correspondance entre le plan de classement Pascal et le plan national de programmation n'a pas pu être conçue automatiquement. De précédents travaux ont montré la possibilité de créer automatiquement une mise en relation entre plusieurs plans de classements (Faucompré 1997). Cette technique nécessite d'avoir un vocabulaire assez proche dans les définitions des divers plans de classement. Nous avons estimé que les deux plans concernés par notre étude

avaient de fortes divergences de terminologie. Pour obtenir un bon résultat, un lourd travail de numérisation (nous ne possédions que des versions papiers) et de recodification du plan de programmation de la recherche algérienne aurait été nécessaire. Cette solution de mise en correspondance automatique n'a donc pas été retenue.

La mise en correspondance a donc été effectuée par une procédure de recherche systématique de tous les codes du plan de classement de Pascal pouvant se rattacher aux programmes de recherche algériens et à leurs axes. Pour ce rapprochement, il n'était pas possible de se contenter du 3^{ème} niveau de la hiérarchie du plan de classement de Pascal. Tous les niveaux ont été pris en compte pour ce rapprochement. Pour faciliter ce travail laborieux, nous nous sommes aidés d'outils informatiques favorisant la recherche et la navigation dans le plan de classement de l'INIST pour repérer les codes candidats. Le plan de classement Pascal complet a été téléchargé depuis le site web de l'INIST. Toutes les pages HTML téléchargées ont été indexées en texte intégral par le logiciel Altavista Discovery⁶. Ce logiciel a permis d'accélérer la recherche et l'identification des codes correspondant à un programme de recherche ou à un axe en appliquant des requêtes systématiques sur tous les termes pertinents qui le décrivent. Le résultat d'une telle recherche étant présenté dans le navigateur Internet, la navigation par les liens hypertextes proposés dans le page HTML du plan de classement de l'INIST était possible favorisant l'exploration transversale du plan de classement selon les liens définis par l'INIST.

L'analyse systématique des codes candidats a permis d'établir la table de correspondance entre les codes Pascal et les programmes de recherche et axes du plan algérien. Une procédure de création d'un nouveau champ *Programme de recherche* dans chaque référence a été automatisée par le logiciel Infotrans. Ce champ répertorie sous forme codifiée le ou les programmes de recherche auxquels les articles se rattachent. La recodification se présente sous la forme d'un descripteur composé d'un code de 9 caractères symbolisant le programme de recherche suivi du titre de l'axe correspondant.

Le tableau suivant présente, par exemple, la table de correspondances établie pour le programme de recherche *Technologies de l'Information* du plan algérien.

Codes programme de recherche/axe	Codes Pascal correspondants
TECH-INFO/ARCHITECTURE DES SYSTEMES	001B00G05B, 001B00G05D, 001D02A01, 001D02A02, 001D02A04, 001D02A08, 001D03H, 001D03I, 001D03J01, 001D03J02, 001D03J03, 001D03J07, 001D03J08, 001D03J09, 001D03J10, 001D03J11,
TECH-INFO-INTELLIGENCE ARTIFICIELLE	001B00G05M, 001D02C
TECH-INFO/MICRO-ELECTRONIQUE	001D03F
TECH-INFO/TECHNOLOGIES DES LOGICIELS	001A01F, 001B00G05H, 001B00G05K, 001B00G05P, 001B00G05R, 001B00G05T, 001B00G05W, 001D02A05, 001D02A06, 001D02A07, 001D02B, 001D03J04, 001D03J05
TECH-INFO/TELECOMMUNICATIONS	001A01E, 001A01G, 001D04

Pour obtenir une correspondance complète entre le plan de classement Pascal et le plan algérien, un 17^{ème} programme de recherche a été ajouté à ceux fixés par le gouvernement algérien. Une grande partie du plan de classement Pascal faisant référence à des aspects techniques non définis dans le plan algérien le "pseudo" programme de recherche *Ingénierie et technologies* a été rajouté (Annexe 2).

Suite à cette recodification des données brutes, l'image de la recherche algérienne s'en voit totalement modifiée. Le graphe suivant (Figure 2) présente la nouvelle répartition des publications algériennes pour la période 90-99 selon les axes prioritaires définis par le plan national.

⁶ Le logiciel Altavista Discovery utilisé dans cette étude est une vieille version qui n'est plus disponible sur le site web d'Altavista. Cette version 1.0 permet d'indexer le contenu des fichiers présents sur son ordinateur personnel. Puis, à partir d'une interface similaire à celle d'Altavista en ligne, l'utilisateur peut rechercher dans le contenu de ses fichiers avec les mêmes commandes de recherche que sur Altavista en ligne.

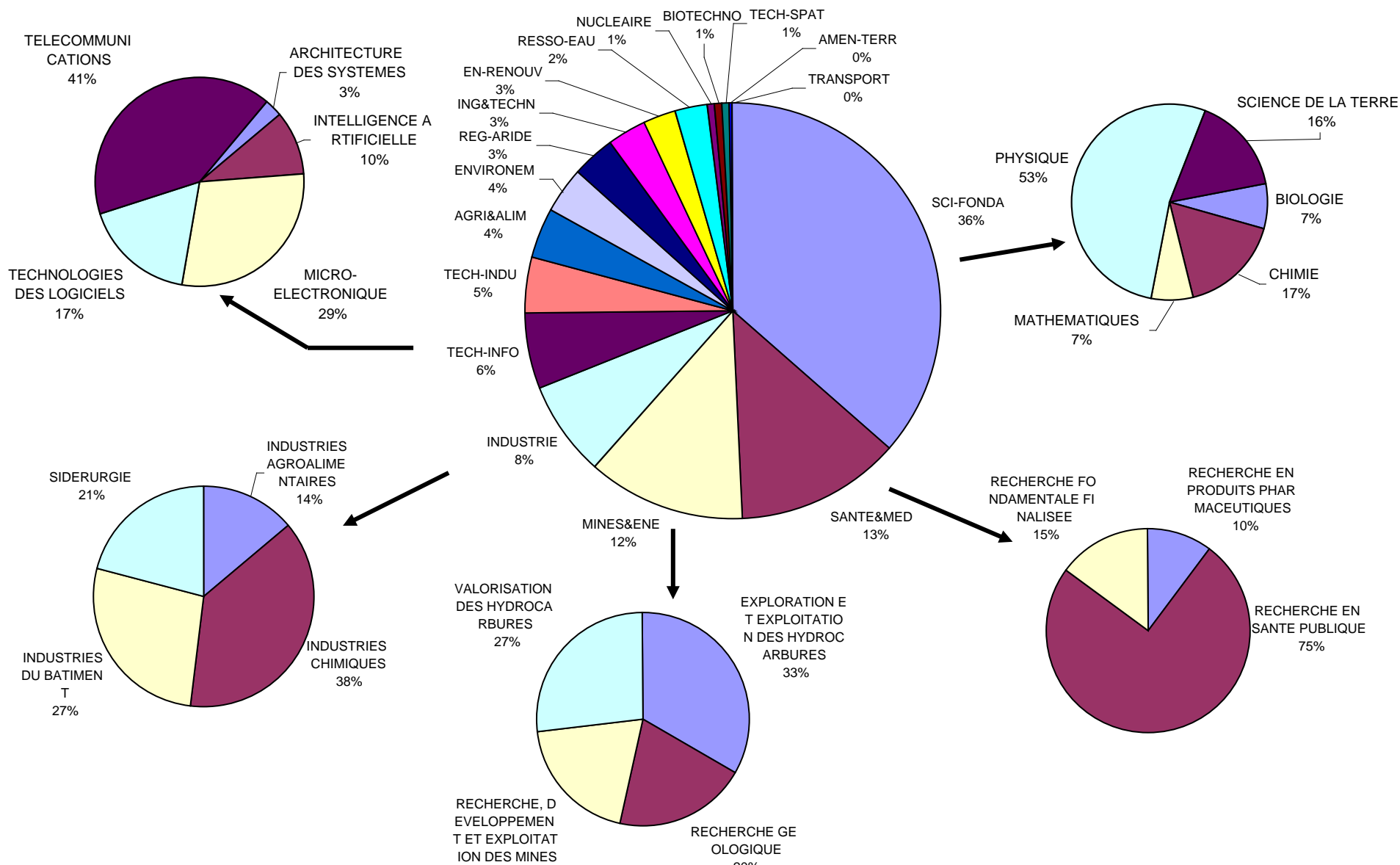


Figure 2: Répartition des publications par programme de recherche algériens

Exploitation bibliométrique des données agrégées

La phase de normalisation et d'agrégation des données terminée, le corpus bibliographique final est alors prêt à l'application d'une grande variété de traitements bibliométriques. Dans le cas de l'étude de l'évaluation de la recherche scientifique algérienne, les résultats bibliométriques sont obtenus à partir des techniques offertes par les logiciels développés au CRRM : Dataview (Rostaing, 1993), Datalist (Pohl, 1997) et Matrisme (Boutin, 1999). Ces traitements bibliométriques permettent de soutenir la réflexion sur l'analyse des potentiels de la recherche scientifique algérienne. La liste suivante expose quelques aspects mis en exergue par ceux-ci :

- Information relatives à l'évolution scientifique au court du temps
 - o Évolution de la production scientifique des institutions et villes algériennes
 - o Évolution des collaborations internationales de l'Algérie
 - o Évolution des pratiques de communication scientifiques algérienne (modification de la politique de publication pour améliorer la visibilité ou une meilleure reconnaissance de la recherche scientifique algérienne)
 - o Évolution des programmes de recherche en Algérie
- Information sur les caractéristiques des villes algériennes
 - o Importance des institutions scientifiques pour chaque ville algérienne
 - o Comparaison des spécialisations des villes algériennes
 - o Affinités des collaborations entretenues par les villes algériennes au niveau local, national et international.
- Information sur la structure de l'organisation de la recherche en Algérie
 - o Réseau de collaborations entre les villes algériennes
 - o Réseau de collaborations entre les institutions algériennes
 - o Réseau de collaborations entre les auteurs algériens (incluant les auteurs étrangers)
 - o Nombre d'institutions algériennes impliquées dans les différents programmes de recherche.

Conclusion

La phase de normalisation et d'agrégation des données brutes, selon des nomenclatures adaptées aux attentes des destinataires, est une phase primordiale dans le processus des traitements bibliométriques. Cette phase est plus délicate à mettre en œuvre lors d'une évaluation nationale que lors d'une action de veille. Cette étape a été la plus longue parmi l'ensemble des traitements réalisés dans cette étude (80% du temps), mais c'est la garantie d'une création de valeurs ajoutées exploitables.

Remerciements

Nous tenons à remercier tout particulièrement Badia Yacine qui nous a aidés dans la réflexion de l'exploitation des résultats bibliométriques pour un usage approprié aux besoins de la politique scientifique Algérienne.

Nous aimerions aussi remercier la société allemande I+K GmbH pour son soutien permanent dans l'exploitation de leur logiciel de reformatage Infotrans.

Bibliographie

BARRE R, "Clustering Research Fields for Macro-Strategic Analysis: A Comparative Specialization Approach.", *Scientometrics*, 22 (1), p.95-112, 1991

BARRE R, LAVILLE F, TEIXEIRA N, ZITT M, "L'observatoire des sciences et des techniques: activités-définition-méthodologie", Solaris, 1995 [<http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2barre.html>]

BOUTIN E, *Le traitement d'une information massive par l'analyse réseau : méthode, outils et applications*, Thèse soutenue à l'Université Aix-Marseille III en 1999

FAUCOMPRES P, *La mise en correspondance automatique de banques de données bibliographiques scientifiques et techniques à l'aide de la classification internationale des brevets*, Thèse soutenue à l'Université de Aix-Marseille III, en 1997

LOI N°98-11, *Journal officiel de la République Algérienne*, 62: 3-42, 1998

NARVAEZ N, RUSSSEL J, ARVANITIS R, WAAST R, GAILLARD J, *La science en Afrique à la fin du XXème siècle. Bilan et perspectives. Impact de la coopération S&T communautaire*. Rapport pour la Commission Européenne DG XII élaboré sous la direction de Waast R (IRD), 1999

POHL L, QUONIAM L, QUAZZOTTI S, DOU H, "Microlevel infometric investigation in phytochemistry", *Journal of chemical information and computer sciences*, 821-827, 1997

Programmes nationaux de recherche, Rapport adopté durant les assises de la recherche scientifique en Juin 1995

ROSTAING H, NIVOL W, QUONIAM L, LA TELA A, "Le logiciel bibliométrique Dataview et son application comme outil d'aide à l'évaluation de la concurrence", Actes du colloque: *Les systèmes d'informations élaborées*, organisé à l'Ile Rousse par la SFBA 9-11 Juin 1993, p. 360-387

Annexe 1 : nomenclature de regroupements de pays

Régions	Pays appartenant à ces régions
AFRIQUE	BENIN, CAMEROON, GUINEA-BISSAU, MALI, NIGERIA, SENEGAL, SOUTH-AFRICA, TANZANIA, UGANDA, ZAMBIA
ASIE	CHINA, HONG KONG, HONG-KONG, INDONESIA, JAPAN, MALAYSIA, PEOPLES-R-CHINA, PHILIPPINES, SOUTH-KOREA, THAILAND
EUROPE DE L'EST	BULGARIA, CROATIA, CZECHOSLOVAKIA, CZECH-REPUBLIC, HUNGARY, POLAND, ROMANIA, RUSSIA, SLOVAKIA, SLOVENIA, UKRAINE, USSR
FRANCE	FRANCE, MARTINIQUE, NEW-CALEDONIA
INDIA/PAKISTAN	INDIA, PAKISTAN
AMERIQUE LATINE	ARGENTINA, BOLIVIA, BRAZIL, CHILE, COLOMBIA, CUBA, JAMAICA, MEXICO, PANAMA, PARAGUAY, VENEZUELA
PAYS DU MAGHREB	LIBYA, MOROCCO, TUNISIA
AMERIQUE DU NORD	CANADA, USA
OCEANIE	AUSTRALIA, NEW-ZEALAND, ZEALAND
PAYS D'ORIENT	ARABIA, EGYPT, EMIRATES, IRAN, ISRAEL, JORDAN, LEBANON, OMAN, SAUDI-ARABIA, SYRIA, YEMEN
EUROPE DE L'OUEST	AUSTRIA, BELGIUM, CYPRUS, DENMARK, ENGLAND, FED-REP-GER, FINLAND, GERMANY, GREECE, ICELAND, IRELAND, ITALY, NETHERLANDS, NORTH-IRELAND, NORWAY, PORTUGAL, SCOTLAND, SPAIN, SWEDEN, SWITZERLAND, TURKEY, UK, WALES

Annexe 2 : Les programmes de recherche nationaux algériens

Ingénierie et technologies

ING&TECHN

Les 17 programmes de recherche retenus	
Titre	Code
Agriculture et alimentation	AGRI&ALIM
Développement des régions arides	REG-ARIDE
Biotechnologies	BIOTECHNO
Environnement	ENVIRONEM
Sciences fondamentales	SCI-FONDA
Santé et médecine	SANTE&MED
Ressources hydriques	RESSO-EAU
Technologies industrielles	TECH-INDU
Industries	INDUSTRIE
Technologies de l'information	TECH-INFO
Mines et énergie	MINES&ENE
Energie et techniques nucléaire	NUCLEAIRE
Aménagement du territoire	AMEN-TERR
Energie renouvelable	EN-RENOUV
Technologies spatiales	TECH-SPAT
Transport	TRANSPORT
Nouveau thème créé	

Programmes de recherche non-retenus
Habitat
Education et formation
Jeunesse et sport
Langue nationale
Traduction
Culture
Economie
Histoire, préhistoire et archéologie
Droit et justice
Population et société
Sciences social
Communication
Langues

Annexe 3 : Nombre de publications pour les codes de 3^{ème} niveau dans la hiérarchie

Catégories des domaines scientifiques	Freq		
SCIENCES DE LA TERRE	438	GENIE MECANIQUE. CONSTRUCTION MECANIQUE	14
ETAT CONDENSE: STRUCTURE ELECTRONIQUE, PROPRIETES ELECTRIQUES, MAGNETIQUES ET OPTIQUES	195	ASTRONOMIE	14
ENERGIE	105	VECTEURS D'IMPORTANCE MEDICALE, NUISANCES, DEPREDATEURS DES DENREES ET MATERIAUX : SURVEILLANCE DES POPULATIONS ET LUTTE	13
ETAT CONDENSE: STRUCTURE, PROPRIETES MECANIQUES ET THERMIQUES	102	RECHERCHE OPERATIONNELLE. GESTION	13
CHIMIE GENERALE ET CHIMIE PHYSIQUE	92	TECHNIQUES D'EXPLORATION ET DE DIAGNOSTIC (GENERALITES)	12
METAUX. METALLURGIE	88	GENIE CHIMIQUE	12
ECOLOGIE ANIMALE, VEGETALE ET MICROBIENNE	75	CHIRURGIE (GENERALITES)	12
ELECTRONIQUE	72	PHYSIQUE ATOMIQUE ET MOLECULAIRE	11
AGRONOMIE. SCIENCES DU SOL ET PRODUCTIONS VEGETALES	72	PHYSIQUE NUCLEAIRE	10
PHYSICOCHIMIE DES POLYMERES	69	ENDOCRINOPATHIES	10
BATIMENT. TRAVAUX PUBLICS	69	GASTROENTEROLOGIE. FOIE. PANCREAS. ABDOMEN	9
MATHEMATIQUES	68	VERTEBRES: ZOOLOGIE GENERALE, MORPHOLOGIE, PHYLOGENESE, SYSTEMATIQUE, CYTOGENETIQUE, REPARTITION GEOGRAPHIQUE	8
SYSTEMES	66	VERTEBRES: REPRODUCTION	8
INFORMATIQUE	66	PRODUCTIONS ANIMALES	8
AUTOMATIQUE THEORIQUE	66	OTORHINOLARYNGOLOGIE. STOMATOLOGIE	8
PATHOLOGIE INFECTIEUSE	65	CYTOLOGIE, MORPHOLOGIE, SYSTEMATIQUE, FLORISTIQUE ET EVOLUTION DES VEGETAUX	8
DOMAINES CLASSIQUES DE LA PHYSIQUE (Y COMPRIS LES APPLICATIONS)	62	PNEUMOLOGIE	7
TELECOMMUNICATIONS ET THEORIE DE L'INFORMATION	60	GYNECOLOGIE. ANDROLOGIE. OBSTETRIQUE	7
RHEOLOGIE	57	BIOCHIMIE ANALYTIQUE, STRUCTURALE ET METABOLIQUE	7
DOMAINES INTERDISCIPLINAIRES: SCIENCE DES MATERIAUX	57	VERTEBRES: ANATOMIE ET PHYSIOLOGIE, ORGANISME DANS SON ENSEMBLE OU ETUDE DE PLUSIEURS ORGANES OU SYSTEMES	6
CHIMIE ANALYTIQUE	50	RADIOTHERAPIE. TRAITEMENT INSTRUMENTAL. PHYSIOTHERAPIE. REEDUCATION. READAPTATION, ORTHOPHONIE, CRENOTHERAPIE. TRAITEMENT DIETETIQUE ET TRAITEMENTS DIVERS (GENERALITES)	6
POLLUTION	48	TOXICOLOGIE	5
ELECTROTECHNIQUE. ELECTROENERGETIQUE	41	PATHOLOGIE OSTEOARTICULAIRE	5
PHARMACOLOGIE. TRAITEMENTS MEDICAMENTEUX	38	MALADIES METABOLIQUES	5
INDUSTRIE DES POLYMERES, PEINTURES, BOIS	38	IMMUNOPATHOLOGIE	5
PHYSIQUE GENERALITES	36	GENETIQUE DES EUKARYOTES. EVOLUTION BIOLOGIQUE ET MOLECULAIRE	5
INDUSTRIES CHIMIQUE ET PARACHIMIQUE	36	TRAUMATISMES. MALADIES DUES AUX AGENTS PHYSIQUES	4
SCIENCES DE L'INFORMATION. DOCUMENTATION	35	SYLVICULTURE	4
PHYTOPATHOLOGIE. ZOOLOGIE AGRICOLE. PROTECTION DES CULTURES ET DES FORETS	32	PSYCHOLOGIE. PSYCHOPHYSIOLOGIE	4
INDUSTRIES AGROALIMENTAIRES	32	OPHTALMOLOGIE	4
CHIMIE ORGANIQUE	26	NEPHROLOGIE. MALADIES DES VOIES URINAIRES	4
SANTE PUBLIQUE. HYGIENE-MEDICINE DU TRAVAIL	25	INFORMATIQUE, STATISTIQUE ET MODELISATIONS BIOMEDICALES	4
INVERTEBRES	25	GEOPHYSIQUE EXTERNE	4
PHYSIQUE DES PARTICULES ELEMENTAIRES ET CHAMPS	24	ANESTHESIE. REANIMATION. TRANSFUSION	4
PHYSIQUE DES GAZ, DES PLASMAS ET DES DECHARGES ELECTRIQUES	22	VERTEBRES: SYSTEME NERVEUX ET ORGANES DES SENS	3
PSYCHOPATHOLOGIE. PSYCHIATRIE	21	TRANSPORTS TERRESTRES, TRANSPORTS AERIENS, TRANSPORTS MARITIMES, CONSTRUCTIONS NAVALES	3
MICROBIOLOGIE	21	IMMUNOLOGIE FONDAMENTALE	3
CHIMIE MINERALE ET ORIGINE DE LA VIE	20	ETHOLOGIE ANIMALE	3
NEUROLOGIE	18	CARDIOLOGIE. APPAREIL CIRCULATOIRE	3
HEMOPATHIES	17		
PHYSIOLOGIE VEGETALE ET DEVELOPPEMENT	16		
DERMATOLOGIE	16		
BIOTECHNOLOGIE. BIODETERIORATION	16		

Annexe 4 : Correspondances entre la nomenclature du CHI et des codes du 3^{ème} niveau du plan de classement de Pascal

001A01	INGENERIE ET TECHNOLOGIE
001A02	MATHEMATIQUES
001A03	INGENERIE ET TECHNOLOGIE
001B00	PHYSIQUE
001B01	PHYSIQUE
001B02	PHYSIQUE
001B03	PHYSIQUE
001B04	PHYSIQUE
001B05	PHYSIQUE
001B06	PHYSIQUE
001B07	PHYSIQUE
001B08	PHYSIQUE
001B09	PHYSIQUE
001B10	PHYSIQUE
001B11	PHYSIQUE
001B12	PHYSIQUE
001B13	PHYSIQUE
001B20	PHYSIQUE
001B30	PHYSIQUE
001B40	PHYSIQUE
001B50	PHYSIQUE
001B60	PHYSIQUE
001B70	PHYSIQUE
001B80	PHYSIQUE
001C01	CHIMIE
001C02	CHIMIE
001C03	CHIMIE
001C04	CHIMIE
001D00	INGENERIE ET TECHNOLOGIE
001D01	INGENERIE ET TECHNOLOGIE
001D02	INGENERIE ET TECHNOLOGIE
001D03	INGENERIE ET TECHNOLOGIE
001D04	INGENERIE ET TECHNOLOGIE
001D05	INGENERIE ET TECHNOLOGIE
001D06	INGENERIE ET TECHNOLOGIE
001D07	INGENERIE ET TECHNOLOGIE
001D08	INGENERIE ET TECHNOLOGIE
001D09	CHIMIE
001D10	INGENERIE ET TECHNOLOGIE
001D11	INGENERIE ET TECHNOLOGIE
001D12	INGENERIE ET TECHNOLOGIE
001D13	INGENERIE ET TECHNOLOGIE
001D14	INGENERIE ET TECHNOLOGIE
001D15	INGENERIE ET TECHNOLOGIE
001D16	TERRE ET ESPACE
001D17	INGENERIE ET TECHNOLOGIE
001E01	TERRE ET ESPACE
001E02	TERRE ET ESPACE
001E03	TERRE ET ESPACE
002A01	BIOLOGIE
002A02	BIOMEDECINE
002A03	BIOMEDECINE
002A04	BIOMEDECINE
002A05	BIOMEDECINE
002A06	BIOMEDECINE
002A07	BIOMEDECINE
002A08	BIOMEDECINE
002A09	BIOLOGIE
002A10	BIOLOGIE
002A11	BIOLOGIE
002A12	BIOLOGIE
002A13	BIOLOGIE
002A14	BIOLOGIE
002A15	BIOLOGIE
002A16	BIOMEDECINE
002A17	MEDECINE CLINIQUE
002A18	MEDECINE CLINIQUE
002A19	MEDECINE CLINIQUE
002A20	MEDECINE CLINIQUE
002A21	BIOMEDECINE
002A22	BIOMEDECINE
002A23	BIOMEDECINE
002A24	BIOMEDECINE
002A25	BIOMEDECINE
002A26	MEDECINE CLINIQUE
002A27	MEDECINE CLINIQUE
002A28	MEDECINE CLINIQUE
002A29	BIOMEDECINE
002A30	BIOMEDECINE
002A31	INGENERIE ET TECHNOLOGIE
002A32	BIOLOGIE
002A33	BIOLOGIE
002A34	BIOLOGIE
002A35	BIOLOGIE
002A36	BIOLOGIE
002A37	BIOLOGIE
002B01	MEDECINE CLINIQUE
002B02	MEDECINE CLINIQUE
002B03	MEDECINE CLINIQUE
002B04	MEDECINE CLINIQUE
002B05	MEDECINE CLINIQUE
002B06	MEDECINE CLINIQUE
002B07	MEDECINE CLINIQUE
002B08	MEDECINE CLINIQUE
002B09	MEDECINE CLINIQUE
002B10	MEDECINE CLINIQUE
002B11	MEDECINE CLINIQUE
002B12	MEDECINE CLINIQUE
002B13	MEDECINE CLINIQUE
002B14	MEDECINE CLINIQUE
002B15	MEDECINE CLINIQUE
002B16	MEDECINE CLINIQUE
002B17	MEDECINE CLINIQUE
002B18	MEDECINE CLINIQUE
002B19	MEDECINE CLINIQUE
002B20	MEDECINE CLINIQUE
002B21	MEDECINE CLINIQUE
002B22	MEDECINE CLINIQUE
002B23	MEDECINE CLINIQUE
002B24	MEDECINE CLINIQUE
002B25	MEDECINE CLINIQUE
002B26	MEDECINE CLINIQUE
002B27	MEDECINE CLINIQUE
002B28	INGENERIE ET TECHNOLOGIE
002B29	INGENERIE ET TECHNOLOGIE
002B30	SANTE PUBLIQUE. HYGIENE- MEDECINE DU TRAVAIL
002B31	ETHIQUE