



**HAL**  
open science

## Les méthodes bibliométriques en soutien d'une approche expert dans la détection de technologies prometteuses

Ivana Roche, Claire François, Dominique Besagni

### ► To cite this version:

Ivana Roche, Claire François, Dominique Besagni. Les méthodes bibliométriques en soutien d'une approche expert dans la détection de technologies prometteuses. VSST 2007, Oct 2007, Marrakech, Maroc. pp.1-12. hal-00310972

**HAL Id: hal-00310972**

**<https://hal.science/hal-00310972>**

Submitted on 25 Sep 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Les méthodes bibliométriques en soutien d'une approche expert dans la détection de technologies prometteuses

Ivana Roche (1), Claire François (1), Dominique Besagni (1)

[ivana.roche@inist.fr](mailto:ivana.roche@inist.fr), [claire.francois@inist.fr](mailto:claire.francois@inist.fr), [dominique.besagni@inist.fr](mailto:dominique.besagni@inist.fr)

(1) INIST - CNRS, 2 allée du Parc de Brabois, CS 10310, 54519 Vandoeuvre-les-Nancy cedex  
FRANCE

---

**Mots-clés :** technologies prometteuses, bibliométrie, approche experts, plan de classement, indicateur de croissance

**Keywords:** promising technologies, bibliometrics, expert approach, classification scheme, growth indicator

**Palabras clave:** tecnologías prometedoras, bibliometría, solicitud experta, plan de clasificación, indicador de crecimiento

**Résumé :** Ce travail s'inscrit dans le cadre d'un projet européen dont le but est l'identification, d'une part, des technologies émergentes et prometteuses ayant des liens forts avec le domaine de la Physique et, d'autre part, des communautés scientifiques développant ces technologies. Nous avons choisi de travailler à partir d'une base de données bibliographique et de nous appuyer sur son plan de classement pour analyser la dynamique des différents domaines thématiques de la base. Cette approche a été combinée à une analyse réalisée par des experts qui nous ont apporté leur vision prospective de leurs domaines de compétence. Dans cet article, nous présenterons les critères qui nous ont guidés pour le choix de la base de données, puis nous décrirons notre méthodologie qui alterne les approches bibliométriques basées sur la notion de plan de classement et les approches par experts pour sélectionner les thématiques prometteuses. Cette sélection est réalisée par étapes successives partant d'une centaine de thématiques, puis les limitant à une quarantaine pour enfin aboutir à 10 technologies prometteuses. L'évolution d'une de ces thématiques est décrite de façon plus approfondie. Nous terminerons par une discussion sur les contraintes associées à ce type de projet.

# 1 Introduction

Plusieurs analyses économiques montrent que l'innovation technologique contribue considérablement à la croissance économique. Cet effet a augmenté cette dernière décennie. De nouvelles technologies comme la microélectronique, les technologies de l'information ou les biotechnologies ont induit l'émergence, ou au moins une croissance importante, de secteurs industriels, et ont changé le profil des pays industriels. Il y a donc un intérêt grandissant pour l'identification de champs technologiques qui auront une exploitation économique à plus ou moins long terme.

En effet, l'analyse des technologies émergentes actuelles montre qu'elles trouvent de plus en plus leur fondement dans les connaissances développées en Physique, le potentiel de contribution de ce domaine scientifique étant considérable. Des recherches récentes en céramiques ont montré l'intérêt de leur utilisation dans la fabrication de capteurs [1] et le développement des micro- et nano-systèmes est largement fondé sur la Physique [2], [3]. Un autre exemple de l'importance grandissante de la Physique dans le processus d'innovation technologique est le traitement de surface avec des caractéristiques très spécifiques qui peut être réalisé par déposition de vapeur physique (PVD) [4]. Ces références montrent que l'émergence des technologies de pointe est chaque fois plus liée aux résultats nouvellement obtenus en Physique, mais on peut également trouver des applications à l'intersection de la Physique et des sciences de la vie. Par exemple, beaucoup de travaux font état des recherches portant sur l'utilisation de matériel biologique comme support pour le stockage de données [5], [6].

Partant de ce constat et l'érigeant en hypothèse, nous avons participé à l'organisation du projet européen PROMTECH en nous définissant les objectifs suivants :

- l'identification des technologies émergentes et prometteuses ayant des liens forts avec le domaine de la Physique ;
- l'identification des communautés scientifiques développant ces technologies ;
- l'obtention d'un outil d'aide à la sélection de thématiques qui seront soutenues dans les programmes de R&D du 7ème Programme-Cadre de la Commission Européenne.

Les travaux que nous présentons s'inscrivent dans le cadre de ce projet.

Les différentes méthodes utilisées pour détecter ces nouvelles tendances, sont en particulier les méthodes Delphi basées sur la confrontation d'avis d'experts et les méthodes de scénarios qui visent à balayer les différents futurs possibles [7]. Une approche complémentaire est l'étude des bases de publications scientifiques et de brevets pour détecter des nouvelles thématiques par analyse bibliométrique. Les méthodologies d'analyse des évolutions se multiplient dans les domaines de la bibliométrie [8] et de la veille technologique [9]. Le programme de recherche TDT « Topic Detection & Tracking » lancé par la DARPA en 1996 a induit le développement de technologies de détection d'événements [10]. Toutes se concentrent sur l'efficacité informatique pour traiter des flux de dizaines ou centaines de milliers de dépêches d'agences, ou d'enregistrements issus d'entrepôts de données ou d'Internet. Pour permettre une analyse des évolutions détectées, des techniques de visualisation de l'évolution sont développées : l'outil ThemeRiver utilise la métaphore de la rivière pour montrer les fluctuations des fréquences de termes dans le temps [11] ; l'outil TCRIP (Temporal Graph Drawing with Intelligent Placement) [12] montre les évolutions de catégories de documents par une superposition de réseaux de catégories calculés à intervalles réguliers. La lisibilité des résultats reste difficile dès que l'on essaie de visualiser des évolutions de structure.

Pour notre projet, nous avons choisi de travailler à partir des notices d'une base de données bibliographique et de nous appuyer sur son plan de classement pour analyser les dynamiques des différentes thématiques recherchées. Nous nous sommes tout particulièrement attachés à la recherche des thématiques présentant un développement stable et fiable dans le temps. Cette approche est couplée à une analyse réalisée par des experts qui nous apportent leur vision prospective de leurs

domaines de compétences. Dans cet article, nous présenterons les critères qui nous ont guidés pour le choix de la base de données, puis nous décrirons notre méthodologie qui fait appel à la fois aux techniques d'analyse bibliométrique basées sur la notion de plan de classement, et aux approches par experts pour sélectionner les thématiques prometteuses. Cette sélection est réalisée par étapes successives, partant d'une centaine de thématiques, les limitant dans un premier temps à une quarantaine pour enfin aboutir à 10 technologies prometteuses.. Une de ces thématiques est décrite de façon plus approfondie en combinant une analyse infométrique et une analyse d'experts.

## 2 Choix des données

L'information nécessaire à notre étude se retrouve soit dans les publications scientifiques, soit dans les brevets. Pour appliquer une méthodologie de type bibliométrique, il est nécessaire d'utiliser une base de données documentaire où l'information est structurée et organisée par l'intermédiaire d'une indexation ou d'un plan de classement de l'ensemble des textes. De plus, cette base doit être suffisamment pluri-disciplinaire pour couvrir les domaines de la Physique, des Sciences Appliquées et des Sciences de la Vie.

A première vue, les brevets semblent les documents les plus appropriés pour analyser les technologies émergentes. Cependant, les bases de brevets ne disposent pas d'un schéma de classification adéquat permettant de faire le lien entre les technologies décrites dans les brevets et les connaissances scientifiques ayant concouru à leur développement. De plus, il était important de pouvoir observer au plus tôt les transferts des connaissances en Physique vers leurs éventuelles applications. Les publications scientifiques, quant à elles, offrent un important potentiel pour repérer les avancées technologiques très en amont, en particulier dans le domaine des technologies basées sur les connaissances. En effet, il est commun de trouver, associés dans la même publication, des résultats innovants et leurs fondements scientifiques. Le choix des membres du projet s'est donc porté sur les publications scientifiques et, plus précisément, sur leurs signalements dans les bases de données bibliographiques.

Un premier problème, lorsque l'on considère les bases de données bibliographiques, est le risque de ne disposer que d'un plan de classement peu détaillé. Cette caractéristique a pour conséquence de limiter les possibilités d'analyse à des interrogations portant sur des hauts niveaux d'agrégation ne permettant pas d'identifier le développement de champs spécifiques. De plus, certaines de ces bases ne proposent qu'une classification unique au niveau de chaque périodique et automatiquement affectée à l'ensemble de ses articles. D'autre part, rappelons que la plupart de ces bases sont spécialisées sur un domaine (Sciences de la Vie, Physique, Chimie...) ce qui interdit d'identifier des liens entre les domaines sources des technologies et leurs applications dans des domaines variés.

Nous avons finalement choisi la base PASCAL de l'INIST, base multidisciplinaire d'environ 15 millions de notices bibliographiques, pour ses caractéristiques, qui répondent à l'ensemble de nos contraintes initiales de travail :

- sa multidisciplinarité, qui permet d'accéder à la fois aux domaines de la Physique et à ceux des applications technologiques ;
- la finesse de son plan de classement, qui offre la possibilité d'analyser des domaines très spécialisés. Par exemple, il est beaucoup plus fin que le «Category codes» du Science Citation Index ;
- son mode de constitution, qui permet d'établir, via les codes de classement multiples, des passerelles entre la Physique et ses applications technologiques. En effet, dans PASCAL, chaque notice bibliographique est affectée d'un ou plusieurs codes et ce de façon individuelle.

Après analyse du plan de classement de PASCAL, nous avons décidé d'employer une stratégie de recherche simple, consistant en la sélection des notices bibliographiques ayant à la fois un code en Physique, quelque soit la spécialité, et un code correspondant à un domaine d'application technologique. Les deux domaines applicatifs sélectionnés ont été : les Sciences de l'Ingénieur et les Sciences de la Vie, qui incluent à la fois les Sciences Biologiques et la Médecine.

## **3 Sélection des thématiques prometteuses**

### **3.1 Analyse des facteurs d'irrégularité**

Afin de pouvoir interpréter les évolutions, il est important de s'assurer que la dynamique des corpus étudiés est bien due à une évolution thématique indépendante de tout événement lié au mode de constitution de la base de données ou à la périodicité de parution de l'information scientifique.

Une analyse des différents événements ayant eu un impact sur la production de la base PASCAL au cours des quinze dernières années, a permis de déterminer l'intervalle de temps ainsi que les types de publications scientifiques que nous devons considérer dans notre étude..

Le début et la fin de deux importantes collaborations d'achat de notices bibliographiques dans les domaines de la Physique et des Sciences de l'Ingénieur nous a conduits à choisir la période entre 1994 et 2003 où ces collaborations étaient actives et stables.

D'autre part, à partir de 2004, la procédure de constitution des notices de la base PASCAL a été en partie automatisée, ce qui conduisit à affecter, dans certains domaines scientifiques, des codes plus génériques à l'ensemble d'une revue et non plus notice par notice comme c'était le cas auparavant. La sélection des notices à partir de l'année de production 2004 est donc rendue plus difficile.

Finalement, nous avons conservé uniquement les articles publiés directement dans les revues, en négligeant les articles publiés dans des actes de congrès, mais également ceux initialement publiés dans des actes de congrès et repris ultérieurement dans une revue. En effet, d'une part, le rythme de parution des conférences est par nature irrégulier ou obéit à une fréquence pluriannuelle et, d'autre part, le nombre d'articles publiés fluctue d'une conférence à l'autre. Ce type de publication induit donc des variations irrégulières de volumétrie qui rendent difficile l'explication des courbes d'évolution de la production scientifique.

### **3.2 Croisement des catégories du plan de classement : sélection du premier ensemble de thématiques**

Le plan de classement de la base PASCAL est une structure hiérarchique constituée de catégories nommées par des chaînes alphanumériques de la forme NNNLNNLNNL..., où N désigne un nombre et L une lettre.

Nous avons exploré les associations entre la Physique (001B), considérée dans sa globalité, et les domaines applicatifs : Sciences de l'Ingénieur (001D) et Sciences de la Vie (002). Pour chacun de ces derniers nous avons défini un niveau dans l'arborescence du plan de classement afin de déterminer la liste des différentes thématiques appliquées devant être examinées. Pour les Sciences de l'Ingénieur (001D), nous avons découpé l'arborescence au niveau « 7 caractères » et nous obtenons par exemple les catégories suivantes :

- 001D03F : électronique des semi-conducteurs

- 001D11C : transformation de matériaux métalliques

Pour les Sciences de la Vie (002), nous avons découpé l'arborescence au niveau « 6 caractères » et nous obtenons par exemple les catégories suivantes :

- 002A04 : biologie moléculaire et cellulaire
- 002B27 : anesthésie, réanimation, transfusion

Notre stratégie de recherche a consisté à sélectionner les références bibliographiques possédant à la fois un code de classement en Physique et un autre dans un domaine d'application. Nous avons ainsi 201 catégories : 133 en Sciences de l'Ingénieur et 68 en Sciences de la Vie. Pour chaque catégorie, nous avons ensuite étudié l'évolution annuelle de sa productivité.

### 3.3 Sélection des 40 thématiques candidates

Cette sélection est réalisée principalement selon des critères statistiques avec une validation finale de la part des experts du Fraunhofer ISI [13].

La première phase de la sélection est fondée sur l'examen du nombre de références de la dernière année de la période considérée. Ainsi, toute thématique ayant, en 2003, un nombre de publications inférieur à 30 a été éliminée. En effet, nous étions tout particulièrement intéressés par des thématiques présentant un développement stable et fiable dans le temps et un nombre trop faible de références ne nous permettrait pas de faire un traitement statistique. D'autre part, les catégories ayant plus de 500 notices en 2003 ont été éclatées en descendant à un niveau plus fin dans la hiérarchie du plan de classement du point de vue applicatif afin de distinguer plus finement les différentes technologies ou domaines scientifiques applicatifs.

Au terme de cette étape, nous avons obtenu 122 catégories dont 14 « petites », c'est-à-dire ayant un effectif en 2003 compris entre 30 et 50 notices. Toutes ont été alors analysées statistiquement, en calculant les indicateurs de croissance suivants :

- l'indice de croissance est une relation simple entre le nombre de publications dans la période la plus récente (2003) et une année de base considérée comme la référence. Deux années de base ont été considérées : 1996 et 2000. Cet indice ne tient pas compte des variations annuelles entre les deux bornes de l'intervalle de temps ;
- le taux moyen de croissance annuelle permet de tenir compte des variations annuelles dans la croissance d'une thématique et complète ainsi l'indice de croissance qui ne tient compte que des bornes de l'intervalle ;
- le « Sharp Ratio » complète le taux moyen de croissance annuelle en le pondérant par la variabilité des croissances annuelles.

Chacun de ces indicateurs apporte une information complémentaire et aucun ne prévaut sur les autres. Aussi nous avons utilisé l'ensemble pour sélectionner les thématiques candidates. La sélection est donc effectuée, d'une part, sur la base de la combinaison de ces indicateurs et, d'autre part, sur l'observation des courbes de croissance, en particulier pour les 14 « petites » thématiques. Par exemple, considérons une thématique n'ayant aucune publication dans les 5 premières années de la période considérée et enregistrant un nombre significatif et croissant de publications de 2001 à 2003. Nous ne pouvons pas calculer ces indicateurs alors que la courbe de croissance est très prometteuse : cette thématique mérite donc d'être sélectionnée. Enfin, cette sélection a été finalisée par les experts qui ont observé le contenu des corpus associés aux thématiques. Ceci a permis, en particulier, d'écarter aussi bien certaines thématiques trop générales, que celles constituées par la concaténation de plusieurs sous-thématiques singulièrement hétérogènes.

Ces étapes successives ont permis in fine de distinguer 45 thématiques candidates que nous avons soumises à un panel d'experts par le biais d'une enquête. Nous avons sélectionné les 10 thématiques prometteuses en analysant les réponses de cette enquête.

### 3.4 Sélection et analyse des 10 thématiques prometteuses

Le panel d'experts a été défini en repérant les principaux auteurs européens dans le corpus associé à chacune des 45 thématiques sélectionnées ci-dessus. Pour cette enquête, nous avons fourni à chaque expert un document contenant un descriptif succinct de chaque thématique le concernant, accompagné d'un ensemble de questions lui permettant de nous faire part de son estimation des potentialités de la thématique.

Les questions peuvent être classées en cinq groupes :

- évaluation des développements futurs, attendus ou pas, de la recherche dans la thématique,
- évaluation des applications futures et de leur impact économique ou dans d'autres secteurs,
- description des obstacles et des facteurs facilitateurs dans les recherches et les applications futures,
- description des développements récents,
- données personnelles concernant l'expert répondant au questionnaire.

Une analyse des réponses a permis de dégager les 10 thématiques prometteuses [14] présentées dans le tableau 1.

Thématique	nombre de références	
	1996-1999	2000-2003
Medical imaging	723	1725
Magneto-electronics	28	253
Molecular biology	1680	2603
Light emitting diodes	1797	2074
Air pollution measurement	173	357
Molecular electronics	53	915
Semiconductors	5241	6112
Surface treatment of metals	93	298
Simulation in mechanical engineering	764	1502
Using neural networks	334	631

Tableau 1. Liste des 10 thématiques finales

## 4 Analyse d'une thématique prometteuse

Une analyse infométrique plus fine des thématiques sélectionnées nous a semblé indispensable pour établir une base de discussion avec les experts de la Commission Européenne lors du séminaire qui clôturera le projet. En effet, cette analyse a permis, d'une part, de décrire précisément chaque thématique et, d'autre part, d'avoir une meilleure appréhension de son évolution dans la période considérée dans l'étude.

Nous présenterons dans cette partie l'analyse réalisée pour la thématique : « Light emitting diodes ».

	Période 2			Période 1		
	MC	Doc.	Intitulé de la classe	MC	Doc.	Intitulé de la classe
<b>Classe 1</b>	20	90	Light emitting devices	9	303	Gallium Arsenides
<b>Classe 2</b>	34	209	Organic semiconductors	33	47	Quantum well
<b>Classe 3</b>	9	376	Photodetectors	9	392	Photodetectors
<b>Classe 4</b>	9	408	Gallium arsenides	20	107	Optical interconnections
<b>Classe 5</b>	21	221	Electroluminescent devices	2	288	Instrumentation
<b>Classe 6</b>	49	54	Semiconductor lasers	19	133	Elemental semiconductors
<b>Classe 7</b>	74	68	Semiconductor materials	29	159	CCD image sensors
<b>Classe 8</b>	24	23	Contact resistance	11	268	Gallium compounds
<b>Classe 9</b>	58	55	Bolometers	42	126	High-speed optical techniques
<b>Classe 10</b>	20	92	Optoelectronic devices	57	51	Laser diodes
<b>Classe 11</b>	69	80	Photodetector	41	37	Superconducting microwave devices
<b>Classe 12</b>	10	549	Light emitting diodes	50	40	Photoconducting devices
<b>Classe 13</b>	16	138	Elemental semiconductors	32	62	Infrared imaging
<b>Classe 14</b>	18	144	Organic light emitting diodes	26	187	Electroluminescent devices
<b>Classe 15</b>	24	103	Cadmium compounds	21	80	Optoelectronic devices
<b>Classe 16</b>	47	95	Pixel	13	352	Light emitting diodes
<b>Classe 17</b>	10	394	Gallium compounds	7	144	Measuring methods
<b>Classe 18</b>	10	467	Electroluminescence			
<b>Classe 19</b>	40	72	Infrared imaging			
<b>Classe 20</b>	13	347	Instrumentation			

Tableau 2. Classes des deux périodes de la thématique « Light emitting diodes », en jaune les classes homonymes

Notre approche est de découper le corpus en deux périodes (1996-1999) et (2000-2003), de réaliser pour chacune une classification automatique en utilisant l'indexation par mots-clés présente dans les notices bibliographiques, et enfin d'analyser l'évolution du vocabulaire décrivant les classes de chaque période. L'outil de classification employé, NEURODOC, applique la méthode de classification des k-means axiales, inspirée du formalisme neuronal des cartes auto-adaptatives de Kohonen comme algorithme de classification non hiérarchique, suivie d'une analyse en composantes principales (ACP) pour la représentation des classes obtenues sur une carte ([15], [16]). Pour analyser l'évolution du vocabulaire décrivant les classes des deux périodes, nous avons construit une matrice de comparaison des mots-clés associés aux classes de chaque période en nous basant sur le pourcentage des mots-clés d'une classe de la seconde période qui préexiste dans une classe de la première période. A partir de cette matrice, nous pouvons alors repérer différents comportements de classes : stabilité, mais également fusion ou éclatement. L'analyse conjointe de cette matrice et des cartes obtenues pour chaque période permet également de déceler des éventuels changements de statut dans le réseau global des classes. Ces différentes observations doivent ensuite être validées par les ingénieurs documentalistes spécialistes de chaque domaine technologique qui réalisent les analyses en étudiant la description complète des classes et des cartes obtenues sous la station d'analyse de l'information STANALYST® [17]. Ces analyses sont ensuite confrontées aux commentaires apportés par les experts externes dans l'enquête décrite ci-dessus.



Notre approche diachronique produit 17 classes pour la première période et 20 pour la seconde. Dans le tableau 2, sont présentés les intitulés des classes de chaque période, ainsi que le nombre de leurs mots-clés et documents.

La matrice de comparaison des mots-clés, présentée dans le tableau 3, montre en ligne, les 20 classes de la seconde période et, en colonne, les 17 classes de la première période. Ainsi, si nous considérons la classe 12 de la seconde période (ligne 12) nous pouvons voir que 10%, 20%, 20% et 40% de ses mots-clés étaient déjà présents dans les classes 8, 10, 14 et 16, respectivement, de la première période. Au total, 90% de ses mots-clés étaient présents dans au moins une classe de la première période. Notons que cette valeur cumule les présences des mots-clés dans différentes classes, elle peut donc dépasser 100%.

21\	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Somme
1	0	10	0	0	0	5	0	5	0	5	0	0	0	25	0	25	0	75
2	2	0	2	0	0	2	0	2	0	0	0	8	0	11	0	11	0	38
3	66	0	44	0	0	0	0	22	11	11	0	0	11	0	0	0	0	165
4	55	0	33	0	0	0	0	44	0	11	0	0	11	0	0	0	0	154
5	0	4	0	0	0	4	0	4	0	0	0	4	0	52	0	14	0	82
6	0	2	0	8	0	0	0	0	4	16	2	4	0	0	2	0	0	38
7	2	9	2	0	0	0	0	0	2	9	0	6	1	2	0	0	0	33
8	4	0	4	0	0	0	0	4	0	8	0	4	0	0	4	4	0	32
9	1	0	1	1	1	0	3	0	5	1	37	8	8	0	1	0	0	67
10	0	0	0	25	5	0	25	0	0	5	0	0	20	0	20	0	0	100
11	0	17	1	0	0	0	0	0	7	0	5	0	0	0	0	0	0	30
12	0	0	0	0	0	0	0	10	0	20	0	0	0	20	0	40	0	90
13	6	6	6	6	0	50	0	6	0	0	0	0	0	0	0	0	0	80
14	0	5	0	0	0	5	0	5	0	0	0	5	0	38	0	38	0	96
15	8	0	12	4	4	0	4	4	4	4	4	12	4	4	4	0	0	72
16	0	8	0	2	0	0	6	0	0	0	0	4	0	0	4	0	0	24
17	40	0	20	0	0	0	0	70	0	10	0	0	0	0	0	10	0	150
18	0	10	0	0	0	10	0	20	0	10	0	10	0	60	0	90	0	210
19	0	0	0	5	2	0	12	0	0	0	0	2	37	0	2	0	0	60
20	7	0	7	7	7	0	30	0	30	15	0	0	15	0	7	0	7	132

classe de 2<sup>e</sup> période provenant d'une unique classe de première période avec même intitulé

classe de 2<sup>e</sup> période provenant d'une unique classe de première période changement d'intitulé

classe dont le vocabulaire n'est pas présent dans la première période

Tableau 3. Matrice de comparaison des mots-clés des deux périodes de la thématique « Light emitting diodes »

L'analyse des lignes de la matrice permet de repérer les classes de la seconde période qui proviennent de classes uniques en première période qu'elles soient homonymes ou non. Il s'agit des classes 5 « Electroluminescent devices », 13 « Elemental semiconductors », 19 « Infrared imaging » et 9 « Bolometers », 11 « Photodetector » respectivement. Les mots-clés de la classe 12 « Light emitting diodes » proviennent majoritairement de sa classe homonyme (16), mais également des classes 10 « Laser diodes » et 14 « Electroluminescent devices » de la première période. Au contraire, l'analyse de la colonne marginale permet de repérer les classes dont le vocabulaire n'est que faiblement présent dans les classes de la première période, et cette particularité nous autorise à émettre une hypothèse de nouveauté pour ces classes.

En analysant les colonnes de cette matrice, nous observons que les classes 14 « Electroluminescent devices » et 16 « Light emitting diodes » de la première période semblent « alimenter », dans la seconde période, leurs classes homonymes (5 et 12) mais également et avec des taux importants les classes 1 « Light emitting devices », 2 « Organic semiconductors », 14 « Organic light emitting diodes » et 18 « Electroluminescence ».

L'analyse des cartes des classes obtenues pour les deux périodes (figure 1), confirme cette observation. En effet, la paire de classes « Electroluminescent devices » et « Light emitting diodes » évolue d'un positionnement marginal dans la première période vers un rôle central dans le réseau de la seconde période, où elles sont associées aux classes qu'elles ont « alimentées ». Dans ce réseau, nous remarquons les classes 14 « Organic light emitting diodes » et 2 « Organic semiconductors » qui portent sur des sujets « nouveaux » trouvant leur origine dans le domaine de l'électronique organique. Cette position dans le réseau central confirme l'importance grandissante de ces nouvelles technologies.

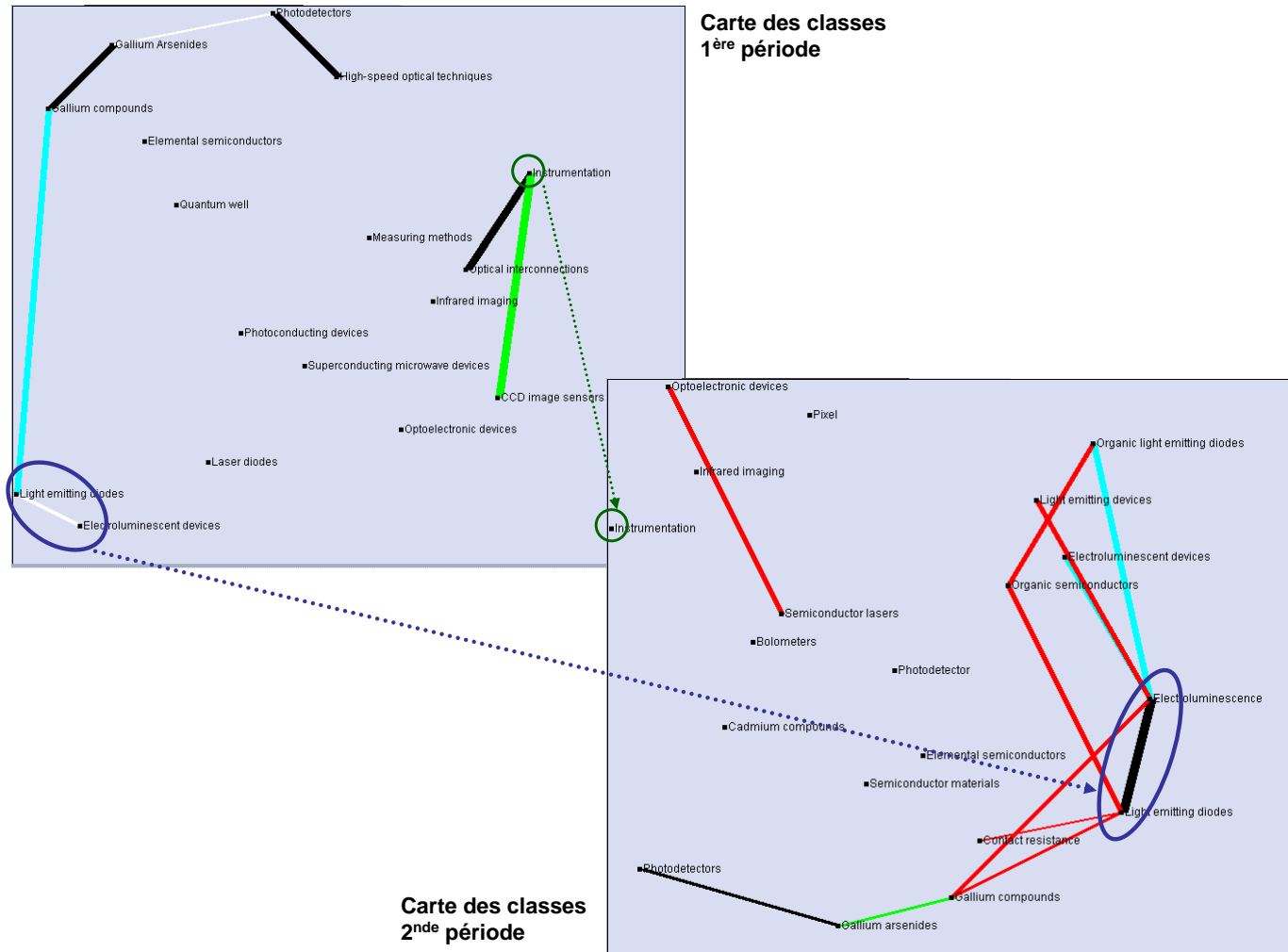


Figure 1. Cartes des classes des deux périodes de la thématique « Light emitting diodes »

Si la thématique générale de ce corpus porte sur dispositifs optoélectroniques à base de semi-conducteurs minéraux ou organiques, en regardant plus précisément au niveau des applications, l'ingénieur documentaliste a pu, par l'analyse des cartes, observer que :

- dans la première période, les applications de type « photodétecteur » (imagerie, capteurs, instruments de mesure, ...) sont prépondérantes, les diodes électroluminescentes sont relativement isolées du reste du réseau des classes
- en deuxième période les diodes électroluminescentes dominent et se retrouvent au centre du réseau des classes

D'autre part, l'analyse du contenu des classes lui a permis de constater, au niveau des matériaux :

- un intérêt croissant de la recherche pour les semi-conducteurs organiques, justifié entre autres par leur souplesse qui permet d'envisager la réalisation de dispositifs électroniques miniaturisés et flexibles ; (voir apparition des classes «Organic light emitting diodes » et «Organic semiconductors» en seconde période ;
- une montée en puissance des nanotechnologies en électronique dans un but de miniaturisation toujours plus poussée des circuits ; (évolution interne des classes «Photodetectors » et « Gallium arsenides ») ;
- un regain d'intérêt pour les alliages semi-conducteurs de type SiGe et pour les semi-conducteurs de type III-V entraîné par l'expansion de la recherche sur les diodes électroluminescentes, déjà observée au niveau des applications (évolution interne de la classe « Elemental semiconductors »)

Finalement, l'analyse des vocabulaires de la classe « Infrared imaging », présente dans les deux périodes, a permis de mettre en évidence un signal très intéressant. Dans la première période, la technologie de l'imagerie infrarouge est associée essentiellement à des applications militaires, tandis que dans la seconde période on observe son association, avec une fréquence significative, avec des mots-clés venant de la biologie et de la géo-détection. Cette observation a permis à l'ingénieur documentaliste d'émettre l'hypothèse soit de l'existence d'un processus de transfert technologique des applications militaires vers les applications civiles, soit, a minima, d'une extension significative en cours du champ d'application de l'imagerie infrarouge.

Nous avons également pu confronter notre analyse aux avis émis par les experts extérieurs qui ont répondu à notre questionnaire (section 1.4). L'examen des réponses obtenues pour cette thématique « Light emitting diodes » font état notamment :

- de l'importance des résultats de la recherche en science des matériaux qui a permis les progrès déjà réalisés dans la thématique et qui permettront ses développements futurs
- des récents développements : les diodes électroluminescentes organiques, les diodes électroluminescentes dans l'ultraviolet profond, les écrans organiques, les lasers quantum-cascade dans le TeraHertz, le photovoltaïque organique
- des applications déjà disponibles : le stockage de données, l'imagerie thermique, la cryptographie quantique, l'éclairage LED, les capteurs bio-médicaux
- des applications futures : les systèmes d'imagerie infrarouge, l'électronique flexible, l'électronique organique
- des champs d'application à l'horizon 2020 : l'optique, les TIC, la défense, la santé, la vie privé
- des impacts sur la société, la qualité de la vie, la sécurité, l'énergie, la santé, les transports
- des facteurs facilitateurs : le financement de programmes de recherche, la coopération entre le monde académique et les entreprises, l'amélioration des infrastructures R&D

Nous pouvons remarquer des points de convergence avec notre analyse, en particulier, dans le signalement des récents développements et des applications futures qui correspondent bien aux évolutions que nous avons observé dans notre analyse diachronique. Par ailleurs, cette phase d'expertise a permis d'enrichir notre connaissance sur le comportement de la thématique avec des informations que nos outils bibliométriques ne sont pas capables de produire, comme les facteurs facilitateurs, les impacts sur la société ou les champs d'application à l'horizon 2020.

## 5 Discussion

Dans ce travail, nous avons utilisé conjointement des techniques bibliométriques et l'analyse des experts. Cette approche nous a permis de réduire la tâche d'expertise en calculant des propositions sur lesquels les experts se sont prononcés. Cette expertise est nécessaire pour valider et compléter les résultats que nous avons pu obtenir avec des méthodes bibliométriques.

Notre démarche bibliométrique est basée sur l'existence d'un plan de classement multidisciplinaire très fin et sur la possibilité d'une indexation des notices bibliographiques par un ou plusieurs codes de classement. Cette contrainte est très forte, et n'est respectée que partiellement. En effet, les méthodes d'indexation sont différentes d'un domaine à l'autre et encore différentes entre l'INIST et ses coopérants. La connaissance précise de ces pratiques a donc été nécessaire pour réaliser l'interprétation des résultats bibliométriques.

De même, les procédures automatiques d'indexation introduites progressivement pour produire la base de données ne permettent plus de respecter cette contrainte. Afin de reproduire ou généraliser cette approche, basée sur un plan de classement, il sera nécessaire d'envisager une méthode de catégorisation automatique des notices dans ce plan et de prévoir un protocole de mise à jour de ce dernier. La difficulté majeure sera la prise en compte du nombre de niveaux et de la finesse de la hiérarchie.

Une approche complémentaire de classification automatique est également à envisager, et plus particulièrement sous une forme incrémentale permettant de suivre les évolutions des classes en fonction de la date de publication des références. Ce type d'algorithme est en cours d'élaboration [18] et des expérimentations portant sur un domaine scientifique unique, bien délimité, ont déjà été réalisées. Cependant, pour analyser une base de données dont la couverture est multi-domaines, nous devons soit considérer une approche de classification à plusieurs niveaux pour définir des classes de plus en plus spécifiques, soit combiner une approche classification avec une méthode de catégorisation automatique.

## 6 Remerciements

Ces travaux ont été réalisés dans le cadre du projet européen « PROMTECH » (PROMisingTECHnologies – Contract N° 15615). Ce projet s'inscrit dans le cadre de l'Action Spécifique NEST (New and Emerging Science and Technology) du 6<sup>ème</sup> Programme-Cadre de l'Union Européenne. Le consortium est constitué par l'ARC System research GmbH (Vienne, Autriche), le Fraunhofer Institut für Systemtechnik und Innovationsforschung (Karlsruhe, Allemagne) et l'INIST-CNRS.

Nous remercions également nos collègues ingénieurs documentalistes de l'INIST-CNRS, et en particulier Nathalie Vedovotto, qui ont participé activement aux différentes étapes du projet en nous apportant leurs expertises scientifiques et documentaires.

## 7 Bibliographie

- [1] LU Z. G. et al., *Recent research and development on semi-conductor gas sensitive ceramics*, Rare Metal Materials and Engineering, 2004, vol. 33, N° 6, supp [1], pp. 14-17
- [2] ROSEI F., *Nanostructured surfaces: challenges and frontiers in nanotechnology*, Journal of Physics Condensed Matter, 2004, vol. 16, N° 17, pp. S1373-S1436
- [3] RAZZACKI S. Z. et al., *Integrated microsystems for controlled drug delivery*, Advanced Drug Delivery Reviews, 2004, vol. 56, N° 2, pp. 185-198
- [4] MATTHEWS A., *Plasma-based physical vapor deposition surface engineering processes*, Journal of Vacuum Science & Technology, 2003, vol. 21, N° 5, supp. [S], pp. S224-S231
- [5] WU P. et al., *Biological nano-ceramic materials for holographic data storage*, Chemical Physics Letters, 2004, 400, pp. 506-510
- [6] TAKITA A. et al., *Data storage inside biological tissue with femtosecond laser processing*, IEIC Technical Report, vol. 105, N° 222, pp. 9-12
- [7] MONTI R. et ROUBELAT F., *La boîte à outils de prospective stratégique et la prospective de défense : rétrospective et perspectives*, Actes des Entretiens Science & Défense, Paris 1998
- [8] NOYONS E., *Science maps within a science policy context*. In Handbook of Quantitative Science and Technology Research, Eds. Moed H.F., Glänzel W., Schmoch U., Kluwer Academic Publishers, London, 2004, pp 237-255
- [9] LOUBIER E. et al., *Visualisation de l'évolution des informations relationnelles par morphing de graphe*. Dans : Journées Francophones Extraction et Gestion de Connaissances (EGC 2007), Namur, Belgique, 23/01/2007-26/01/2007, [Cépaduès Editions](#), p. 43-54, janvier 2007
- [10] BINSZTOK H. et GALLINARI P., *Un algorithme en ligne pour la détection de nouveauté dans un flux de documents*, Dans : JADT 2002 : 6<sup>èmes</sup> Journées Internationales d'Analyse Statistique des Données Textuelles
- [11] HAVRE S. et al., *ThemeRiver: visualizing thematic changes in large document collections*, IEEE Transactions on Visualization and Computer Graphics, 2002, Vol. 8, N°1
- [12] ERTEN C. et al., *Exploring the computing literature using temporal graph visualization*, Report, Department of Computer Science, University of Arizona, 2003
- [13] FRIETSCH R., *Results of statistical evaluation*, Deliverable 03 for Project PROMTECH (contract N° 15615), 12 pages, 2007
- [14] von OERTZEN J., *Results of evaluation and screening of 40 technologies*, Deliverable 04 for Project PROMTECH (contract N° 15615), 32 pages + appendix, 2007
- [15] LELU A., *Modèles neuronaux pour l'analyse de données documentaires et textuelles*, Thèse de l'Université de Paris 6, 1993
- [16] POLANCO X. et FRANÇOIS C., *Data Clustering and Cluster Mapping or Visualization in Text Processing and Mining*, In: Dynamism and Stability in Knowledge Organization: proceedings of the Sixth international ISKO conférence, 10-13 July 2000, Toronto, Canada. Edited by Clare Beghtol, Lynne C. Howarth, Nancy J. Williamson: Ergon Verlag, 2000, pp 359 - 365
- [17] BESAGNI D. et al., *Stanalyst® : Une station pour l'analyse de l'information*, Dans : Actes de Veille Stratégique Scientifique et Technologique VSST2004, Toulouse 25-29 octobre 2004, pp 319-320
- [18] LELU A. et al., *Classification dynamique d'un flux documentaire : une évaluation statique préalable de l'algorithme GERMEN*, JADT 2006, Besançon, 19-21 Avril 2006