HAL
open science

Identification of emerging technologies through pathway-III analysis

Edgar L. Schiebel, Marianne Hörlesberger, Dirk Holste, Ivana Roche, Dominique Besagni, Claire François

To cite this version:

Edgar L. Schiebel, Marianne Hörlesberger, Dirk Holste, Ivana Roche, Dominique Besagni, et al.. Identification of emerging technologies through pathway-III analysis. IAMOT'2010, 19th International Conference on Management of Technology, Mar 2010, Le Caire, Egypt. hal-00614078

HAL Id: hal-00614078

https://hal.science/hal-00614078

Submitted on 9 Aug 2011

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.

IDENTIFICATION OF EMERGING TECHNOLOGIES THROUGH PATHWAY-III ANALYSIS

EDGAR SCHIEBEL[1)]

[1)] AIT Austrian Institute of Technology GmbH, Tech Gate Vienna, Donau-City-Straße 1, 1220 Vienna, Austria, edgar.schiebel@ait.ac.at

MARIANNE HÖRLESBERGER[1)]

marianne.hoerlesberger@ait.ac.at

Dirk Holste [1)]

dirk.holste@ait.ac.at

IVANA ROCHE [2)]

[2)] INIST - CNRS, 2 allée du Parc de Brabois,
CS 10310, 54519 Vandoeuvre-les-Nancy Cedex, France
ivana.roche@inist.fr

DOMINIQUE BESAGNI [2)],

dominique.besagni@inist.fr

CLAIRE FRANÇOIS [2)]

claire.francois@inist.fr

Abstract

Early recognition of emerging technologies is very important for keeping or gaining a leading role in the markets of companies, for research organisations to stay in the front end of research as well as for national or international funding agencies to promote the right future technologies. We propose advanced methods and algorithms based on bibliometric analysis applied on large-scale databases for scientific publications in order to identify emergent technologies. In the PROM-TECH project we demonstrated in a proof-of-concept approach the possible and reliable detection of growing research activity for a variety of technological fields. Building on this work, in ongoing investigations we analysed in-depth research progress for single technologies defined by classification categories of the PASCAL database.

Using keywords that addressed a more disaggregated level of a technology, however, allows for in-depth analysis of selected technology fields. We constructed a diffusion model of three diffusion stages to establish a framework for the arrangement of keywords according to their history. We statistically estimated the parameters of the diffusion model for two successive time periods of four years and were thus able to model the development of “keyword

pathways". Pathways were defined as the change of a keyword from one stage to the next during subsequent time periods. We showed that technological breakthroughs in technologies could be verified by related keywords changing from stage-1 of the diffusion model to stage-3 (which we termed pathway-III). Hence pathway-III analysis was strongly indicative of technological breakthroughs in emerging technologies.

Here we extend the analysis already conducted for single technological fields to the comprehensive survey of all 45 identified fields in PROM-TECH. We identify and extract keywords related to emerging technologies through a pathway-III analysis and show that it reveals new promising perspectives for emerging technologies.

Co-word maps demonstrate if predefined technology classes reflect a coherent emergence of term combinations or new combinations of terms do emerge cross over predefinitions. We could show that new emerging fields tend to combine research findings cross over many technology fields.

Introduction

An enduring problem in the identification and development-tracking of new and growing research topics, as well as emerging technologies, lies in their proper and fast recognition when lack of expert knowledge and short time windows practically constrains any detailed analyses. Additionally, emergence comes very often stealthy and hardly noticeable, and very often existing research findings are a matter of interdisciplinarity and therefore hardly recognisable by any expert. Advanced methods and algorithms based on bibliometric analysis hold the promise to purposefully exploit large-scale databases for scientific publications in order to identify emergent technologies (Takeda Y. et al., 2009; Shibata N. et al., 2008; Roche I. et al., 2008).

In previously published works, we demonstrated in a proof-of-concept approach the possible and reliable detection of growing research activity for a variety of technological fields (PROM-TECH, 2007), which bundled extracted topics in high aggregated levels like "Optoelectronic devices" or "Semiconductors". Building on this work in ongoing investigations we analysed in-depth research progress for single technologies (Schiebel E. et al., 2009).

Our research has generally shown that the definition of a technological field is tied to the nature of the underlying research and very context-dependent. Because new technological fields are still in a *status nascendi*, where new research findings are increasingly expressed, labelled and diffuse with non-standard terminology within the research community, they are neither comprehensively defined nor extracted by simple retrieval techniques with predefined technology classes. Using keywords that addressed a more disaggregated level of a technology, however, allows for in-depth analysis of selected technology fields. Keywords of publications are matter-of-factly related to effects, methods, materials, devices or applications of the natural sciences. We thereby noticed that such keywords occurred in articles of clearly different technology fields and that they possessed a "historical trait", i.e., they originated from a home technology and diffused to other technology fields. We constructed a diffusion model of three diffusion stages to establish a framework for the arrangement of keywords

according to their history (Roche I., et al., 2008). Parameters of the diffusion model were statistically estimated for two successive time periods and thus a model for the identification of "keyword pathways" was built. Pathways were defined as the change of a keyword from one stage to the next during subsequent time periods. We showed that technological breakthroughs in technologies, e.g., the emission of bright white light of LEDs, could be verified by related keywords changing from stage-1 of the diffusion model to stage-3 (which we termed "pathway-III"). Hence pathway-III analysis was strongly indicative of technological breakthroughs in emerging technologies.

Here we extend the analysis already conducted for single technological fields to the comprehensive survey of all 45 fields identified in PROM-TECH. We extract keywords related to emerging technologies through a pathway-III analysis and show that it reveals new promising perspectives for emerging technologies.

Up to now we were only able to identify growth in pre-classified technology fields or pathways of single terms of predefined technology classes. In this paper we identified all terms of pathway-III for all 45 technology fields and examined their interrelations by their co-occurrence in publications cross over all fields. Co-word maps demonstrate if predefined technology classes reflect a coherent emergence of term combinations or new combinations of terms do emerge beyond established fields.

Data

Data related to the 45 identified fields in PROM-TECH have been extracted from the corresponding classification categories of the PASCAL bibliographic database for two time periods: between 2005 and 2006 (P1) and between 2007 and 2009 (P2). Longer time periods were not possible because of changes in the database production.

The second period includes the years 2007 and 2008 and only partially 2009. We wanted to include the newest publications. As it was not possible to divide a year for some equal 2.5 years periods we took these two unequal periods nevertheless. It was not a methodological problem, because we did not compare absolute quantities for the periods but relative.

For all the 45 considered technological fields, the PASCAL database query delivered 12,712 documents. In the first period we had 4,884 documents and 12,047 different keywords among which 5,032, (41%) appear less than 3 times. The second period had 7,828 articles and 14,306 terms among which 7,180 (50%) occur less than 3 times.

In Table 1, we detail the 45 technology fields, their definition, theirs corresponding PASCAL categories and, finally, the number of articles obtained for each considered time period.

Table 1: Definition of technology fields and number of publications in the two periods

ID	Database classification categories	Technology field	Articles P1: 2005 and 2006	Articles P2: 2007 - 2009

1	001D02D01; 001D02D13	Theoretical automation, general	36	-
2	001D02D09; 001D02D10; 001D02D11; 001D02D12	Process control	109	237
3	001D03E	Electronique tubes	26	61
4	001D03F	Electronics, semi conductors	745	1748
5	001D03F04	Transistors	163	325
6	001D03F12	Magnetics	8	46
7	001D03F15	Optoelectronics	352	695
8	001D03F18	Molecular electronics	81	525
9	001D03F21	MEMS	95	172
10	001D03G02A	Electronic circuits	44	92
11	001D03I	Information storage	53	82
12	001D05A	Electrical engineering, general	1	1
13	001D05C	Electrical materials	7	29
14	001D05D	Electrical machines	8	27
15	001D05G	Electrical elements	21	40
16	001D05I	Electrical energy	19	116
17	001D06D01; 001D06D05; 001D06D06; 001D06D08; 001D06D09; 001D06D10; 001D06D11	Thermal energy, combustibles	53	65
18	001D07O	Cristallisation, lixivation, separation	1	5
19	001D08A02	Minerals		
20	001D08B04; 001D08B05	Ceramics	883	773
21	001D11A	Metalurgy, general	146	109
22	001D11C02	Thermical treatment of metals	89	150
23	001D11C04	Mechanical working of metals	125	215
24	001D11C06	Surface treatment of metals	414	993
25	001D12A	Mechanical engineering, general	371	532
26	001D12E	Mechanical transmissions	456	684
27	001D12F	Mechanical motors	168	231
28	001D14F	Soil mechanics	141	273
29	001D14L	Hydraulic engineering	79	156
30	001D15C	Road traffic	30	60
31	001D15D	Rail traffic	8	30
32	001D16B	Waste treatment	7	9
33	001D16E	Global pollution	12	42
34	002A01	Molecular biology,	55	81

		general		
35	002A02	Biochemistry	95	91
36	002A03	Molecular biophysics	59	43
37	002A04	Cell biology	32	32
38	002A08A	Biophysics of tissues, general	99	337
39	002A08B; 002A08C; 002A08D	Biophysics (materials, magnetics, electrical properties, mechanics)	61	145
40	002A08E; 002A08F	Physical impact on biological material	16	9
41	002A14	Ecology (animals, plants, cells)	150	8
42	002A22	Cardiovascular systems	18	53
43	002A25	Nerves, sense organs	199	153
44	002A31	Biotechnology	84	385
45	002B26	Physical therapy	96	188

Methodology

We used a diffusion model to build a framework for the classification of keywords in different stages of development. The next step was the classification of the keywords according to diffusion stages for the mentioned two periods. The migration of terms from one stage in period 1 to another stage in period two marked so called pathways. These pathways described the career of keywords. The migration of terms from a stage of genesis in period one to a stage of high diffusion in stage two was defined as emergence. As terms described only very singular characteristics of a technology, we had to recombine these terms according to their usage in the publications. Firstly we draw a co-word map with the software BibTechMon™. It revealed the coherence of terms usage related to classified categories on one hand and autonomous coherence by the usage of the authors. We sharpened the picture of coherence by a cluster analysis. We did this in two ways. Firstly the clustering of co-occurrences of pathway-III terms with each other in publications gave us an idea about an emerging technology as a bunch of different characteristics. In a second so called diachronic cluster analysis we combined publications where such pathway-III terms occurred and examined the evolution of the emerging technologies over the two time periods.

Diffusion model

To build the model we start with the idea that technologies have their roots in an invention which was created by one person or a small group of scientists who use specific terms for describing it. After a specific invention is published, the respective knowledge is available to other scientists. If scientists are inspired by the new ideas, they make use of the new theories and technologies, publish them and add their own ideas. At that stage, a new terminology is created and described in few publications. New terms are created or applied coming from other disciplines but get sometimes a new meaning. If the findings are of interest for other

research groups they pick them up and the number of publication grow. In this second stage the findings are primarily interesting for researchers in the same discipline. That is why growth takes place in this area. If the findings are fundamental also for the use in another discipline the related terms diffuse over the scientific community and they occur in publications of many disciplines. Investigating the terminology of a suitable corpus of publications at a certain time span we can divide the terminology into three stages, which expresses an invention, new findings, growth in a discipline and diffusion in the scientific community:

(i.) STAGE 1: New terms created or adapted from other fields in connection with a specific considered technology (or discipline, etc.) can be found in a few publications. These terms occur with a low frequency in this considered field, and they form an exotic cluster.

(ii.) STAGE 2: Already established specific terms in the considered technology (or discipline, etc.) occur in this stage 2. We call the technology field, where a corpus of specific terms is established, the Home Technology for these terms.

(iii.) STAGE 3: This is the stage where we consider the so called cross section terms. Such terms are well established and applied in different technology fields. They are well understood not only in the considered field. They show a broad diffusion to other research or technology fields. One can detect a high number of articles in different technology fields using these terms.

We call this model of the three stages *Diffusion Model*. Therefore, the diffusion model identifies three categories of terms: terms unusual in a considered field, the established terms in a considered field, and cross section terms regarding to a considered field.

To be able to examine diffusion over different technologies we have to assign the so called *Home Technology* to each term. Terms occurring with a higher probability in a considered field than in another field are defined as such Home Technology terms. For defining these terms we used the 45 technology fields identified in the PROMTECH-Project. The probability is defined by the frequency (number of publications) of one term in a technology field divided by the absolute number of articles in this field. This probability is calculated for the same term for all technology fields in our corpus. The technology field with the highest probability is defined as the Home Technology for the considered term.

Secondly, we used a measure for the *Cross Section Degree* of a term in a selected technology. For this purpose, we introduced the Gini Index (Gini Index, 2008), a measure of statistical dispersion most prominently used as a measure of inequality of income distribution or inequality of wealth distribution. The Gini Index varies from 0 to 1. In our case, 0 means a completely uniform distribution of all terms and indicates that each term of the considered field occurs in all other fields of our corpus. A Gini Index of 1 tells us that all terms of the considered field occur only in this field, which means that no term of the considered field occurs in another field. Therefore, more general terms have a lower Gini Index.

The bibliometric interpretation of the diffusion model with three stages allowed to assign each keyword to a diffusion stage: Stage 1: Birth seldom or unusual use of terms; Stage 2: growth of usage of a term in articles of the home technology; Stage 3: high number of articles and diffusion of terms to other technologies (high number of articles in different technologies).

We used a bibliometric filter to classify keywords according to different stages of the diffusion model, see Table 2.

Table 2: Bibliometric filter to classify terms by diffusion stages

Indicator	Selection Criteria	Stages of Diffusion
technology field	all 45	-
minimum number of articles	more than 2 articles	-
broad diffusion: small Gini Index for low concentration of terms in different technology fields	a third of all terms	stage 3 terms
low Diffusion: large Gini Index	the other two thirds	stage 1 or stage 2
high frequency of publications	next third of all terms	stage 2 terms
low frequency of publications	last third of all terms	stage 1 terms

The filter was sensitive to 1) the number of articles; 2) the Gini index as a measure for the unevenness of the probability distribution of keywords over 45 technological fields; and 3) the relative frequency of terms for two time periods. A tracking procedure of keywords through the different stages for two periods revealed pathway-III significant keywords.

Pathways

We describe the dynamics of emerging technologies by pathways of keywords through diffusion stages from one period to the following one. Figure 1 illustrates the pathways. The dynamics of emerging technologies can be revealed by identifying the keywords that follow the mentioned pathways, see Besagni et al. (2009).

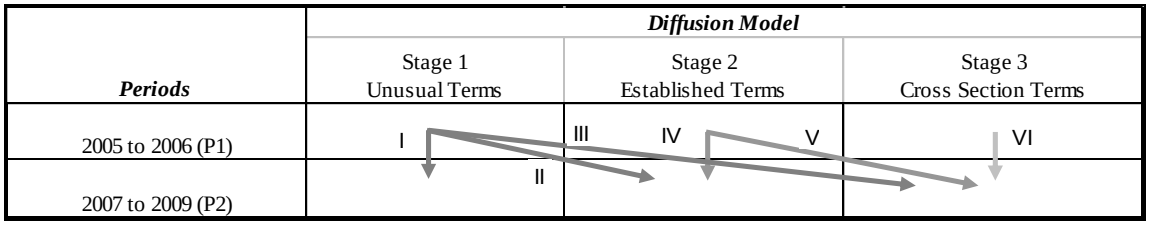


Figure 1. Taxonomy of the migration pathways of keywords (research topics) through different stages of the diffusion model in two time periods, see Schiebel et al 2009

In this paper we concentrate on terms that follow the pathway-III. They are looked upon as keywords with the shortest career of emergence.

Co-word mapping

The interrelation of identified pathway-III terms are examined by a co- word analysis. For the visualisation and analysis we used the software BibTechMon™. The core of the software is an algorithm to position related objects in a two dimensional space. The relation of terms in our case is defined by their pair wise co-occurrence in publications. These co-occurrences are normalized by the often used Jaccard index. The terms are arranged in a network structure by a spring model coming from mass point mechanics. In this way a term is positioned with respect to all other ones. A term in the centre of the network or a sub network has a more or less strong relation ship to many other nodes in the network graph, what is called a high centrality. In our case, established terms or very common terms are centred. New terms and unusual terms are positioned in the periphery. A bunch of terms that are cited often together cross over many publications form up in a sub network.

Diachronic cluster analysis

In a last step we re-incorporate the keywords into their context by a diachronic clustering of corresponding documents (Polanco X., et al. 2001). The diachronic cluster analysis is realized with the help of a clustering tool applying first a non-hierarchical clustering algorithm and then a principal component analysis to map the obtained clusters. This tool is implemented in the information analysis platform Stanalyst®. Neurodoc (Lelu & François, 1992; Lelu, 1993), the clustering tool we used, applies a non-hierarchical clustering algorithm, the axial K-means method, coming from the neuronal formalism of Kohonen's self-organizing maps, followed by a principal component analysis (PCA) in order to represent the obtained clusters on a 2-D map.

Our approach consists of:

(i.) splitting the corpus in two successive time periods, namely 2005-2006 and 2007-2009;

(ii.) applying cluster analysis on the corpus of each period, in which documents are represented by the keywords existing in the bibliographic references;

(iii.) analysing the evolution between the two cluster sets and maps by examining the vocabulary related to the clusters of the two periods.

In order to analyse the evolution of the cluster vocabulary between the two considered periods, we build a comparison matrix pointing out the percent of keywords belonging to the second-period clusters and already existing in the first-period clusters. The cumulated percentage is also calculated for each second-period cluster. Using this matrix, we can identify different cluster behaviours: stability, fusion or splitting. Using the cluster maps, we can also detect status change of the clusters in the global network.

Results

The first step to obtain pathway-III terms was the application of the bibliometric filter to arrange the terms in the stages of diffusion. The distribution of terms by the number of publications reveals a high number of keywords occurring only once or twice. To avoid this

noise we selected terms with at least 3 publications. A first categorisation in stage 3 was done by extracting keywords that occurred in publications that were assigned to many classification categories or to many technologies in our notation. We used a threshold of 0,9642 for the Gini Index in period 1 and 0,95556 in period 2. The criterion was to obtain 3 stages with an equal number of terms. This was arbitrary but pragmatic to construct a framework for a further analysis. The remaining two thirds of terms with a higher Gini Index (low diffusion over other technology fields) were separated by low and high frequency in stage 1 and stage 2, see Table 3.

Table 3. Filter to identify stages of diffusion for terms

Indicator	**Indicator value**	**Number of terms**		**Stages of diffusion**
		P1: 2005-2006	P2: 2007-2009	
technology field	all 45	12,047	14,306	
minimum number of articles	more than 2 articles	5,032	7,180	
broad diffusion: small Gini Index	P1: Gini <0,9642 P2: Gini <0.95556	1,679	2394	stage 3 terms
low Diffusion: large Gini Index	P1: Gini >0,9642 P2: Gini >0.95556	3,353	4,786	
high frequency	P1: freq.>4 articles P2: freq. >4 articles	1,670	2665	stage 2 terms
low frequency	P1: freq.<5 articles P2: freq.<5 articles	1,683	2,121	stage 1 terms

Based on this filter list of terms according to the diffusion stages have been generated in a database. According to the definitions of pathways in Figure 1 the lists of terms were examined on identical terms. The Table 4 shows the number of terms in each pathway. For example we calculated the number of keywords with pathway III by identifying the same terms in the list of stage 1 in the first period and in the list of terms of stage 3 in second period.

Table 4: Number of keywords in different pathways

Pathways	**Number of keywords**	**Percent of terms in P2**
I	299	14.1%
II	459	17.2%
III	276	11.5%
IV	856	32.1%
V	462	19.3%

VI	1228	51.3%

Pathway-III terms per technology

Pathway-III terms demonstrate shift from a new and rather exotic status in the first period to a status of high relevance in the second period for several technological fields. As a first approach we sorted the 45 technologies by the number of pathway-III terms, see Table 5. It is an astonishing fact that "Cardiovascular systems" with only 18 publications in period 1 and 53 publications in period 2 hold the highest number of pathway-III terms. The technology with the second highest number of pathway-III terms is "molecular biology, general", where we have 55 publications in period 1 and 81 publications in period 2. We found the lowest number of pathway-III terms for "ceramics" with one term and "surface treatment of metals" with two ones.

Table 5: Number of pathway-III terms per technology

Technology	Number of pathway-III terms
Cardiovascular systems	17
Molecular biology, general	13
Hydraulic engineering	12
Electrical energy	12
Transistors	12
Thermal energy, combustibles	11
Molecular electronics	11
Biotechnology	10
Magnetics	10
MEMS	9
Electronic tubes	9
Electronic circuits	9
Thermal treatment of metals	9
Optoelectronics	9
Mechanical working of metals	8
Information storage	8
Electrical machines	8
Mechanical engineering, general	8
Road traffic	8
Metallurgy, general	8
Biophysics of tissues, general	6
Soil mechanics	6
Cell biology	6
Electrical materials	6
Mechanical motors	6

Molecular biophysics	6
Physical therapy	5
Mechanical transmissions	5
Electrical elements	5
Process control	4
Biophysics (materials, magnetics, electrical properties, mechanics)	4
Rail traffic	4
Global pollution	3
Nerves, sense organs	3
Biochemistry	3
Surface treatment of metals	2
Ceramics	1

A deeper look at two technologies with the highest number of pathway-III terms is given in the tables 6 and 7. These tables reflect pathway-III terms. We have to consider, that each of these keywords occurred in several technologies but with highest probability in the mentioned one. As we can see in figure 3 the terms of Cardiovascular systems do not form a sub network with a high density, but show very low co-occurrences. This means that there is not a significant combination of these terms in the period 2. The publications reveal that there are many articles dealing with models of blood flow: models with viscous flow in elastic tubes, injured endothelial wall, a rheological model for the inelastic blood rheology with viscoelastic fluid, application of parallel algorithm for the numerical simulation of the blood flow in human brain, coupling of an elastic structure with an incompressible fluid. Central terms in this research also occur in a couple of other technologies, which could be an interesting interdisciplinary combination.

Table 6: Pathway-III terms in Cardiovascular systems

Keyword	Number of articles in P1	Number of articles in P2	Number of technologies
Vibrations	3	22	11
Adhesive joint	3	15	10
Collagen	3	13	7
Simple support	4	13	7
Viscoelastic fluid	3	13	8
Viscous fluids	4	12	8
Euler equation	3	10	5
Numerical stability	3	10	5
Stokes flow	3	10	8
Viscous flow	3	10	7
Fiber orientation	3	9	6
Parallel algorithm	4	9	6
Penalty method	3	9	6
Internal flow	3	8	5
Distributed parameter system	3	7	6

Prostheses	3	7	6
Revolution shell	3	7	4

There are 13 pathway-III terms with Molecular biology as a home technology. Figure 3 together with a study of the relevant publications suggests, that there is no focus on an emerging issue in molecular biology visible by observation with pathway-III terms. The terms are spread over the landscape and different topics are stressed by the articles. The migration of these terms from stage 1 in period P1 to stage 3 in period P2 must have other reasons than a coherent development of a research issue in molecular biology.

Table 7: Pathway-III terms in Molecular biology, general

Keyword	Number of articles in P1	Number of articles in P2	Number of technologies
Photoinduced effect	4	23	8
Scanning microscope	4	15	7
Colloidal suspension	3	14	5
Diffuse reflection	3	13	5
Biotin	3	10	8
pH	4	9	7
Thiols	3	9	6
Ellipsometry	4	8	7
Optical losses	4	8	5
Laser welding	4	6	4
Stochastic equation	3	6	5
Target tracking	3	5	5
Focal length	3	4	4

Co-word map of pathway-III terms

Emerging technologies tend to appear as a bunch of activities that are closely related or recombined in a new manner. A good possibility to show this kind of interdependencies is given by co-word maps.

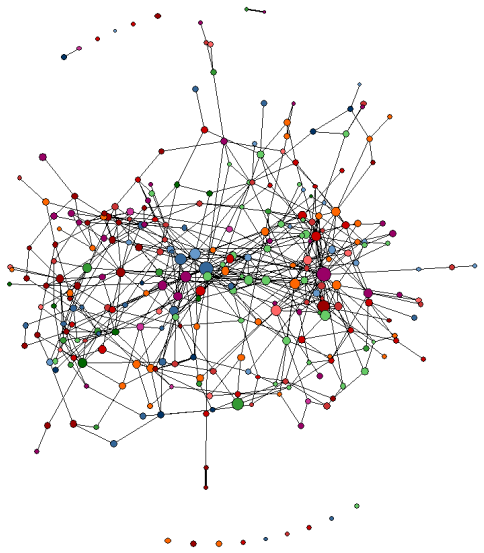
The key question, we wanted to investigate with co word mapping is: Do technologies defined by classification categories reveal emerging new (sub)technologies or do we have much more significant new combinations that form a new emerging technology, that has not been identified before and established by a classification code.

The co-word map of all pathway-III terms is shown in Figure 2. In this picture we demonstrate the interrelationships between the terms with a focus on the technology fields defined by the classification codes.

The map shows a big core of interrelated terms and some peripheral single topics. There is a remarkable interconnection between the terms that could not be expected in this form. We expected more separated sub networks because of the different nature of the selected 45

technologies. It seems that these emerging topics do not represent particular and separated research but reflect high interdisciplinarity in the combination of methods, materials, models, etc.

The 45 technologies are represented by the colours in the map. A deeper view at each of the technologies shows that there is no significance that one of them forms a coherent sub network. “Photovoltaic cell” takes a very prominent part with 46 articles and 25 co-pathway-III terms. Therefore it can be assumed that this is an application with the highest emergence in our investigation. The second one is “Drug dispensation” with 25 articles and 14 co-pathway-III terms.



Number of terms	276
Number of edges	557
Network density	0.015

Actor	Articles	Co-words	Actor Degree Centrality
Photovoltaic cell	46	25	0.091
Drug dispensation	25	14	0.051
Optical characteristic	21	14	0.051
Experimental result	24	13	0.047
Electrophoresis	17	13	0.047
Photoinduced effect	13	13	0.047
Imagery	42	12	0.044
Fourier transform spectroscopy	18	11	0.040
Energy method	18	10	0.036
Vibrations	17	10	0.036
High energy	14	10	0.036
Surface chemistry	12	10	0.036
Quality control	11	10	0.036
Blue light	23	9	0.033
Non invasive method	19	9	0.033
Planar technology	17	9	0.033
Therapy	17	9	0.033
Quartz	16	9	0.033
Adhesive joint	14	9	0.033
Relaxation	14	9	0.033
Circular cylinder	12	9	0.033
Indium	11	9	0.033
Conducting material	10	9	0.033
Double layers	10	9	0.033
Scanning microscope	10	9	0.033
...	...	...	...

Co-word map of pathway 3 terms, colours are related to the 45 home technologies of the terms

Network indicators and selected terms

Co-word map of terms according to their co-occurrence in articles, circle: term, size proportional to number of articles, connection: Jaccard index of co-occurrences, periode: 2007-2009, BibTechMon™

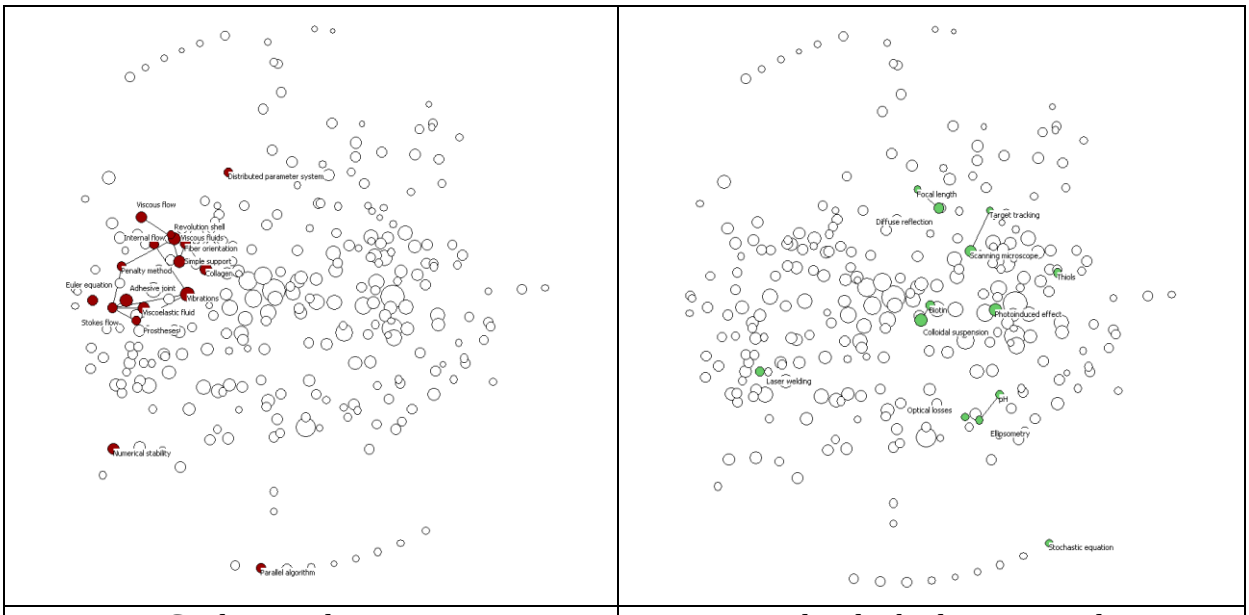
Figure 2: Coword map and network indicators for pathway-III terms

To demonstrate the interrelationship and the (non)coherence of terms of the technologies defined by the classification codes we highlighted the terms with the assigned home technology as shown in Figure 3. As it has already been mentioned before, the co-occurrences of such terms in articles is very weak. It should be expected that emerging technologies tend to have a broad interest of a growing number of researchers and new upcoming terminology is used in publications what usually leads to high co-occurrences between such terms. We could not show this for the predefined technologies.

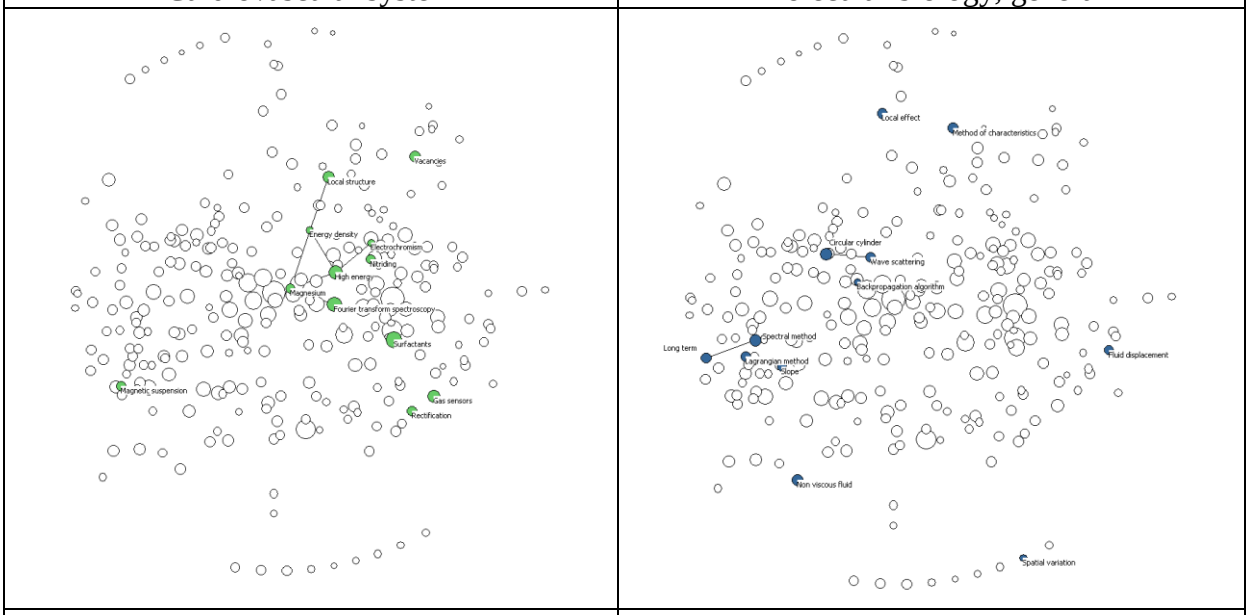
For the technology field “Cardiovascular system” there is a little higher interconnection of the terms by co-occurrences in articles but all other fields have only a few interrelations. On the other hand the network density of the whole map is much higher (see Figure 2) than in the

association of terms by the classified technology fields. This means, that research activities do form other bunches of term combinations.

We obtain a complete different picture, if we identify clusters of terms based on their interrelations as shown in Figure 2.



Cardiovascular system	Molecular biology, general



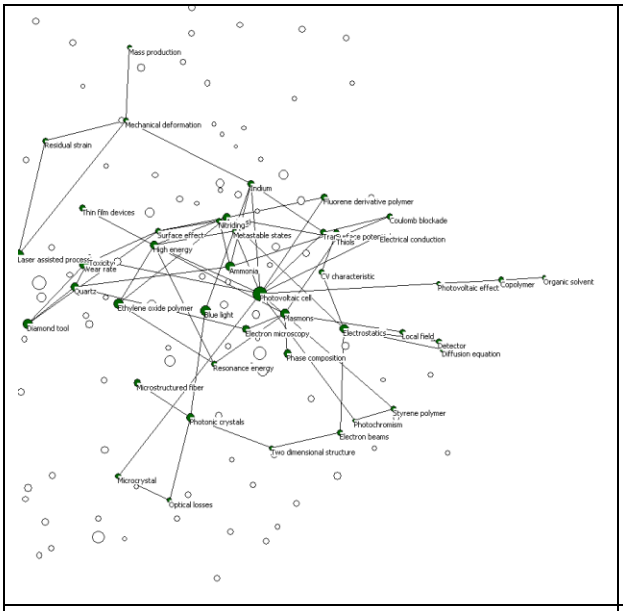
Electrical energy	Hydraulic engineering

Co-word map of terms according to their co-occurrence in articles, circle: term, size proportional to number of articles, connection: Jaccard index of co-occurrences, period: 2007-2009, BibTechMon™

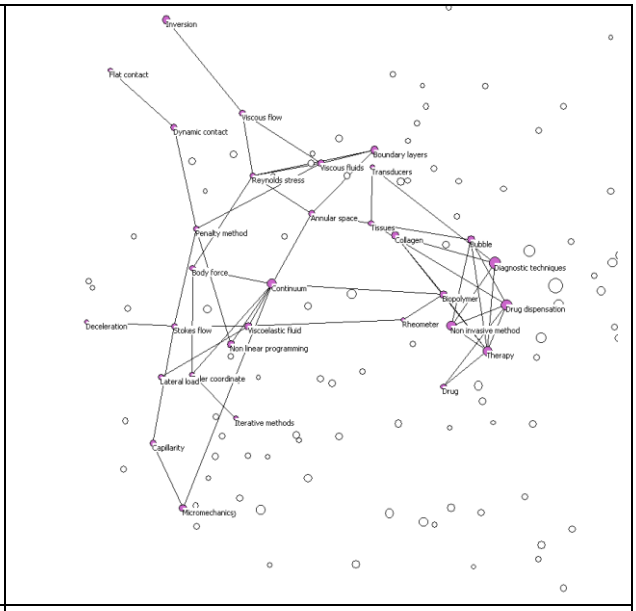
Figure 3: Co-word map of pathway-III terms – Four selected technologies highlighted

The results of a cluster analysis (Pearson correlation and average linkage agglomeration on the basis of co-occurrences) revealed seven compact clusters of pathway-III terms: Photovoltaic cell, Imagery, Laser deposition, Anodizing, Drug dispension and Diagnostic techniques, Planar technology and Optical characteristic as well as Energy method. The names of the clusters are the most frequent representative terms.

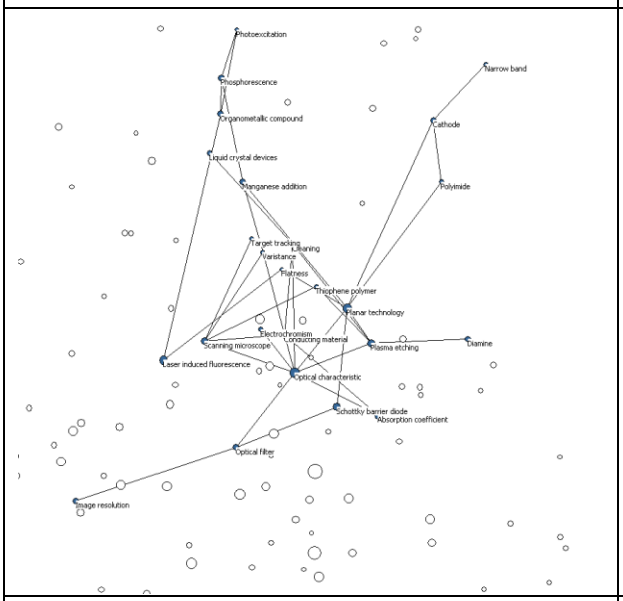
The co-occurrences of the pathway-III terms of four most compact clusters are shown in Figure 4. All four clusters have a much higher interrelationship than the pathway-III terms associated by classified technology fields.



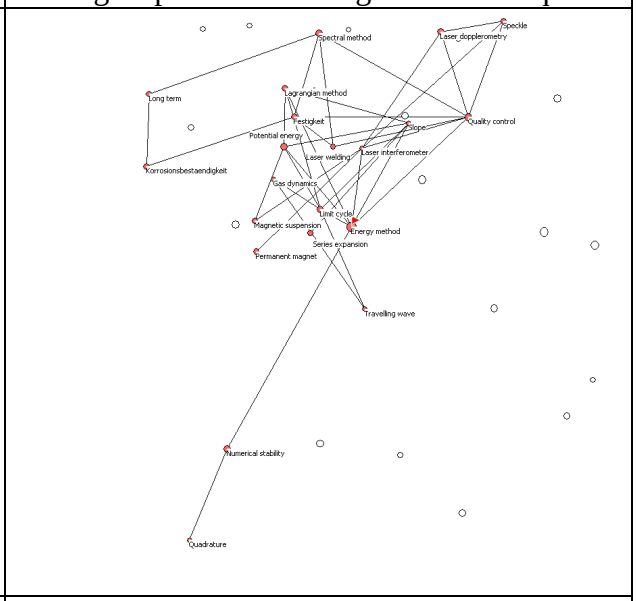
Photovoltaic cell



Drug dispension and diagnostic techniques



Planar technology and optical characteristic



Energy method

Co-word map of terms according to their co-occurrence in articles, circle: term, size proportional to number of articles, connection: Jaccard index of co-occurrences, period: 2007-2009, BibTechMon™

Figure 4: Co-word map of pathway-III terms – Four selected clusters

Diachronic cluster analysis

The diachronic cluster analysis applied on the two periods P1 and P2, allows us to determine which topics of the second period have roots in the first one and which topics are new in the second period. The clustering algorithm produced 20 clusters involving 4,868 documents and 3,650 keywords in period P1, and 20 clusters relating 7,827 documents and 7,398 key-words in period P2. Let us notice that the important reduction of the number of keywords associated to the clusters in both periods, around 35% with respect to the initial number of terms obtained directly from the query results, does not entail a significant loss of information at the level of the documents. Indeed we had 99.7% and 100% of the initial information corpora in the clusters of respectively P1 and P2. We can see in Figures 5 and 6 the obtained cluster maps for the two periods.

Comparing the cluster contents of the two maps we observe, first of all that a great number of

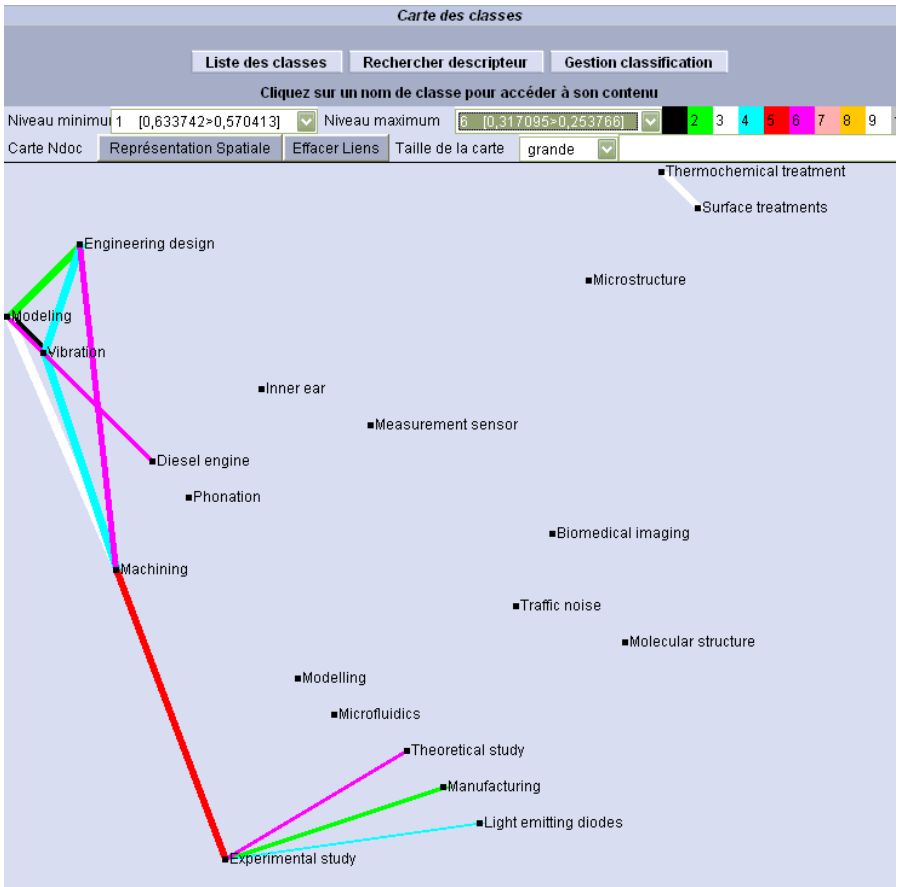


Figure 5: First period cluster map of the 45 technological fields

clusters in the second period remain stable, coming predominantly from their homonym in the first period. They are: “Modelling”, Manufacturing”, “Vibration”, “Biomedical imaging”, “Light emitting diodes”, “Microstructure” and “Microfluidics”. The two others homonymous clusters, “Surface treatments” and “Modelling”, have a different behaviour and are mainly formed by, respectively, one (“Thermo chemical treatment”) and 3 (“Machining”, “Vibration” and “Engineering design”) more clusters, besides the homonymous cluster. These 9 homonymous clusters contribute to structure the networks in both periods.

In the first period we have, in the left side of the map, a great network gathering a set of clusters dealing with mechanical engineering (“Engineering design”, “Modelling”, “Vibrations” and “Diesel engine”) and clusters related to electronics (“Light emitting diodes” and “Experimental study”). The cluster “Machining” links these two sub-networks. In the right upper side is located a little network dealing with chemical engineering and metallurgy, clusters “Thermo chemical treatment”, “Surface treatments” and “Microstructure”. The three clusters related to life sciences research topics (“Biomedical imaging”, “Phonation” and “Inner ear”) have isolated locations in the cluster map.

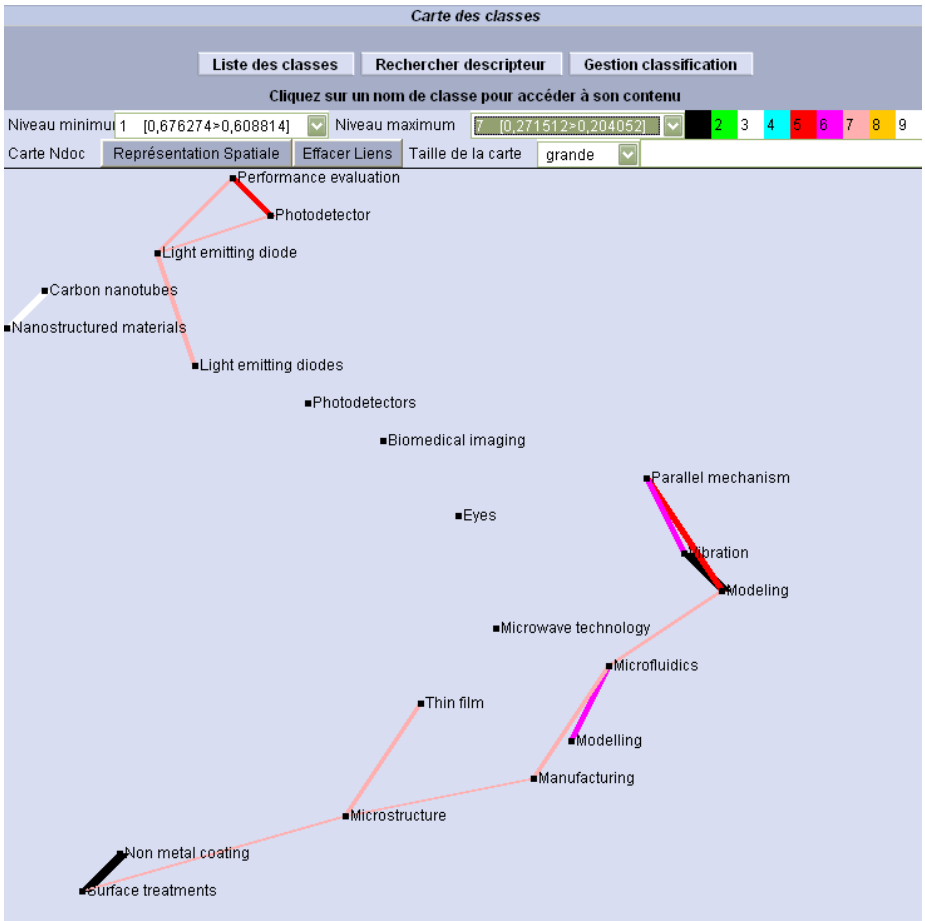


Figure 6: Second period cluster map of the 45 technological fields

In the second period map we find, in the half lower side, a big network corresponding to the restructuring of the two networks observed in the first period map. Indeed the splitting of the

cluster “Machining” contributes to separate the mechanical engineering sub-network from the electronics one. Furthermore it facilitates its grouping with the small network dealing with chemical engineering and metallurgy. The consequence is the formation, in the second period, of this large network starting from mechanical engineering, in its right part, to chemical engineering in its left part going through metallurgy. In the left upper side of the map, appears a network of clusters dealing with electronics and coming from a single cluster (“Light emitting diodes”) in the first period. Indeed, this cluster is at the origin of important new clusters: “Photodetectors”, “Light emitting diode”, “Nanostructured materials” and “Carbon nanotubes”. This evolution shows mainly the increasing of optoelectronic devices and the emergence of new research topics related with nano-technologies. The life sciences are represented by two clusters still in isolated position (“Biomedical imaging” and “Eyes”) but with size increase.

Convergence between diffusion model and diachronic cluster analysis

The comparative analysis of the results of the two approaches operated in this work, namely the diffusion model and the diachronic cluster analysis, reveals some quite interesting convergence points.

We study the location, on the P2 cluster map (Figure 6), of the pathway-III terms related to the 4 technologies given in figure 3 and we compare with the co-word maps presented in Figure 3:

(i.) Cardiovascular system: the 17 pathway-III terms are grouped in the left side of the co-word map and they are present in the 4 clusters of the right part of the biggest network (“Parallel mechanism”, “Vibration”, “Modelling” and “Modeling”). This concentration in clusters related to mechanical engineering shows the need in technical knowledge to manage medical issues, especially in fluid mechanics, concerning the modelling of blood flow.

(ii.) Molecular biology: the 13 pathway-III terms are located in the right side of the co-word map and they are present in the electronics network (mainly “Light emitting diodes”, “Nanostructured materials” and Photodetector”), in the life science clusters (“Biomedical imaging” and “Eyes”) and in the middle of the biggest network (“Manufacturing” and “Thin film”). This fact is reflected in some publications about bio-inspired nanostructured materials.

(iii.) Electrical energy: the 12 pathway-III terms are also in the right side of the co-word map and they are present in the electronics network (mainly “Light emitting diodes”, “Performance evaluation”, “Nanostructured materials” and “Carbon nanotubes”) and in the left part of the biggest network (“Microstructure”, “Surface treatments” and “Manufacturing”).

(iv.) Hydraulic engineering: the 12 pathway-III terms are dispersed in the co-word map with some groupings in its left side. Some of them are present in two clusters (“Vibration” and “Modeling”) located in the right part of the biggest network.

Conclusions

In the PROM-Tech project we developed a method to identify emerging technologies and applied it with the PASCAL bibliographic database. The study of the growth of the number of publications and an expert evaluation generated 45 promising emerging technologies from the point of view of expected research breakthroughs and high economic potential. These technologies had been defined on the basis of the classification categories assigned to each document by the database producer. On this level only highly aggregated already fields were visible and emergence was supposed to happen somewhere inside of this field. We performed a deeper view inside of these fields by analysing diffusion characteristics of terms of document keyword lists. In our first approaches we focused on single fields out of the 45 technologies and examined dynamics of keyword careers within the boundaries of one technology field. We observed that many research activities in the shape of terms were used in a timely dynamic and changing manner in a stable, growing or descending, number of the identified technology fields. We made the obvious assumption that such terms combine to emerging research as a bunch of activities cross over the predefined technologies. We extracted such activities by using a diffusion model over two periods and identified so called pathway-III terms that migrated from a stage of genesis to a stage of high diffusion to many research activities. These terms did not have high co-occurrence in publications when we focused them on their assignment to the predefined technologies. A co-word map revealed an intense interrelationship of the pathway-III terms with each other. Terms with a high centrality were Photovoltaic cell and Drug dispension. They are representatives of two important research areas one in engineering and one in medicine. Significant combinations of pathway-III terms by co-occurrences could be identified with a cluster analysis: Photovoltaic cell, Imagery, Laser deposition, Anodizing, Drug dispension and Diagnostic techniques, Planar technology and Optical characteristic as well as Energy method.

The evolution of the whole field with all keywords of the 45 technologies, obtained by the diachronic cluster analysis, showed still high dynamics in research about both, well known emerging technologies with, for example,the increasing of optoelectronic devices, and the emergence of new research topics related with nano-technologies. The life sciences are represented by two clusters (“Biomedical imaging” and “Eyes”) still in isolated position in the map but with an increase of their size.

The appearance of pathway-III clusters in the whole evolution of the technology fields was studied by their location in the cluster map of period P2. The terms associated with “Cardiovascular system” are grouped in the left side of the co-word map and they are present in the 4 clusters of the right part of the biggest network (“Parallel mechanism”, “Vibration”, “Modelling” and “Modeling”). This concentration in clusters related to mechanical engineering shows the need in technical knowledge to manage medical issues, especially in fluid mechanics, concerning the modelling of blood flow. Pathway-III terms related to Molecular biology were present in the electronics network (mainly “Light emitting diodes”, “Nanostructured materials” and Photodetector”), in the life science clusters (“Biomedical imaging” and “Eyes”) and in the middle of the biggest network (“Manufacturing” and “Thin film”). This fact is reflected in some publications about bio-inspired nanostructured materials.

References

Besagni D., François C., Hörlesberger M., Roche I., Schiebel E. (2009): Les émergences technologiques dans le domaine des dispositifs optoélectroniques : identification et caractérisation, paper presented at the séminaire, VSST (Veille Stratégique Scientifique et Économique), 30 - 31 march 2009, Nancy.

Gini Index (2008): *http://en.wikipedia.org/wiki/Gini_coefficient*

Lelu, A., François C. (1992): Hypertext paradigm in the field of information retrieval: A neural approach, 4th ACM conference on hypertext, November 30th-December 4th, Milano.

Lelu, A. (1993): Modèles neuronaux pour l'analyse de données documentaries et textuelles, PhD Dissertation, Université de Paris VI.

Polanco X., Francois C., Royauté J., Besagni D., Roche I. (2001) STANALYST (2001): An Integrated Environment for Clustering and Mapping Analysis on Science and Technology, 8th International Conference on Scientometrics and Informetrics, July 16-20 2001, Sydney, Australia, Proceedings Vol. 2, pp. 871-873.

PROM-TECH (2007): Identification and Assessment of Promising Emerging Technological Fields in Europe, final report July 2007, AIT, Vienna.

Roche I., Besagni D., Francois C., Hörlesberger M., Schiebel E. (2008): Identification and Characterization of Technological Topics in the Field of Molecular Biology, 10th International Conference on Science and Technology Indicators, Vienna, 17- 20 September 2008, pp. 320-322.

Schiebel E., Hörlesberger M., Roche I., François C., Besagni D. (2009): An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices, paper presented at the 12th International Conference on Scientometrics and Informetrics, ISSI 2009, 14-17 July, 2009, Rio de Janeiro, Brazil.

Schiebel E., Hörlesberger M. (2007). About the Identification of Technology Specific Keywords in Emerging Technologies: The Case of "Magnetoelectronic". 11th ISSI Conference, Madrid, Spain 25-27 June.

Shibata N., Kajikawa Y., Takeda Y., et al. (2008): Detecting emerging research fronts based on topological measures in citation networks of scientific publications, Technovation 28/11/758-775

Takeda Y., Kajikawa Y. (2009): Optics: a bibliometric approach to detect emerging research domains and intellectual bases, Scientometrics 78/3/543-558.