Identification of Emergent Research Issues: the Case of Optoelectronic Devices

Ivana Roche, Nathalie Vedovotto, Dominique Besagni, Claire François, Roger Mounet, Edgar L. Schiebel, Marianne Hörlesberger

▶ **To cite this version:**

HAL Id: hal-00959338
https://hal.science/hal-00959338

Submitted on 14 Mar 2014

Identification of Emergent Research Issues: the Case of Optoelectronic Devices

Ivana Roche[1], Nathalie Vedovotto[1], Dominique Besagni[1],
Claire François[1], Roger Mounet[1],
Edgar Schiebel[2] and Marianne Hörlesberger[2]
[1]Institut de l'Information Scientifique et Technique (INIST-CNRS)
Allée du Parc de Brabois, CS 10310, 54519 Vandœuvre-lès-Nancy
[2]Austrian Institute of Technology (AIT)
Tech Gate Vienna, Donau-City-Straße 1, 1220 Wien
[1]France
[2]Austria

1. Introduction

The optoelectronic devices field is one of the last decade's most promising technological fields. Light emitting diodes gain more applications in cars and housing lighting, OLED displays are introduced in electronic devices and consumer electronics. Optimization of lighting power and tuning of light spectrum are well known important research topics.

Early recognition of new and alternative products and production processes is a strategic necessity contributing to appropriate assessment and decision-making. Emerging technologies are essential to advances in research, industry and society. But the detection of emerging technologies remains unsolved.

An interesting introduction to the use of bibliometrics for the identification of upcoming issues from online databases was published by Lancaster & Lee (1985). These authors analyzed the spread of an issue from pure sciences literature to popular press, through applied sciences literature. They measured the occurrence of a keyword over time in different online databases covering scientific literature. Additionally they suggested a procedure to identify growing issues in terms of time gradients of the number of published articles where single keywords occur.

The selection tree introduced by Armstrong & Green (2007) gives an additional picture of the landscape of the forecasting methods that could be employed to detect these emerging technologies. This tree illustrates the dichotomy between judgmental and quantitative forecasting methods and shows the great diversity of existing approaches like the Delphi or the Nominal Group Technique ones. Forecasts are obtained in a structured way from two or more experts. Other methods combine expert domain knowledge with statistical techniques and allow the identification of causal forces acting on trends.

Even though the number of available data is large enough to apply quantitative methods, the important question of the data type used for forecasting remains. The two data types mostly used to detect new topics with bibliometric analysis are provided by bibliographic

databases covering scientific literature and patent databases. These methods consist of simple statistical techniques, such as growth curve analysis, or more sophisticated ones, such as clustering or network analysis. A third data type, related to state-funded research grants, is very interesting. The analysis of these three sources of information as a whole, and not as separate entities, allows gaining an understanding of the triple-helix interfaces between university, industry and government.

The framework for this study was the PROMTECH project (PROMTECH (2007)), financed by the European Commission. The main goal of the project was to elaborate a methodology enabling the identification of promising emerging technologies.

As the optoelectronic devices field turns out to be among the most promising ones, we focus on it by identifying its emerging research topics. We then study their evolution by considering data sets of bibliographic records related to two successive time periods and represented by their associated keywords. We apply this diachronic approach to the following analytical methodologies:

- a "Diffusion Model" using a bibliometric filter that distributes keywords in different diffusion stages in order to model the field terminology evolution (Schiebel & Hörlesberger 2007). The visualization of the results was operated by the software BibTechMon™, a bibliometric monitoring tool (Kopcsa & Schiebel 1998);

- a cluster analysis enabling the identification of emerging research topics by comparing the clusters related to each period (Roche et al. 2008). The clustering was operated by the software-tool Stanalyst (Polanco et al. 2001) that produces a graphical representation of the clustering results under the form of a network of clusters.

Firstly we refer to the data acquisition. Secondly the applied methodologies are illustrated. Thirdly the results of the applied methodologies to the field ''optoelectronic devices'' are presented. Finally we discuss the two approaches, summarize and conclude our discussion.

2. Data

The data framework is a sample recorded from PASCAL, a multidisciplinary bibliographic database produced by the French institute for the scientific and technological information (INIST-CNRS), that was specifically adapted to the purpose of our approach. PASCAL provides broad multidisciplinary coverage of scientific publications and contains, nowadays, about 20 million bibliographic records that are derived from the analysis of the scientific and technical international literature published predominantly in journals and conference proceedings. PASCAL is appropriate for the intended analysis, as each analyzed document is registered by a very fine classification, much more sophisticated than the category codes often employed in other sources of bibliographic data. The PASCAL classification categories, named also classification codes, belong to a structured classification scheme hugely detailed that is a taxonomy of every field and subfield of all the scientific disciplines covered in the database.

In addition, PASCAL covers international publications in physics, chemistry, engineering sciences, life sciences, and medicine and produces an indexing, on the one hand, by keywords and, on the other hand, by classification categories, both given in the database and assigned to the individual publications, either manually by scientific experts or automatically based on a content analysis. The query operated in this work was based on the information conveyed by these indexations. By extracting from the obtained corpus of bibliographic records the keyword indexing we produced a terminology related to the

studied domain. After a verification step, done by a scientific expert, that terminology can be employed in our analysis.

The search focused on publications related, on the one hand, to Optoelectronics by means of classification codes or keyword indexing and, on the other hand, to peripheral scientific fields whose research issues, even if apparently far from the concerns of the researchers working on Optoelectronics, could have an impact on this domain. The main idea is to produce a finite space of exploration constituted by the information contained in the set of records answering the search strategy. The choice of these peripheral scientific fields is delicate. Indeed, if it is too large, the exploratory space could present a high level of noise with a great number of records that will be definitively uninteresting and that we would need to eliminate before analysis. But conversely, if the operated choice is too sharp, the exploratory space could not reflect in a suitable way the scientific environment richness of the studied domain and this silence would be very harmful for its analysis.

After some iterations, the list of peripheral fields converged towards: "Physics", "Chemical Industry", "Physico-chemistry of polymers" and "Polymer industry". The choice of these fields was motivated by their interest towards the synthesis and characterization of materials involved in optoelectronic devices. A search strategy has been detailed:

- firstly, we looked for the PASCAL classified simultaneously in the field "Optoelectronics" and in a field that is either "Physics" or "Chemical Industry" or "Physico-chemistry of polymers" or "Polymer industry":
- then we added the records indexed by the keywords "*LED*" or "*Electroluminescent device*" or "*Optoelectronic device*" and classified either in "Physics" or "Chemical Industry" or "Physico-chemistry of polymers" or "Polymer industry".

The result is a corpus of 8,169 bibliographic records all in all, divided into the periods from 2000 to 2004 and 2005 to 2010 delivers:

- 2,590 records for the first period (2000-2004)
- 5,219 records for the second period (2005-2009/2010[1])

The 8,169 bibliographic records (all years, 2000-2010) are assigned to more than 160 different PASCAL codes, whereas a reference can be assigned to more than one code.

The first 20 fields with the biggest number of records in the two considered periods (2000-2004 and 2005-2010) are showed, respectively, in Tables 2 and 3. Among the 16 fields matching in both periods, 12 confirm their position or rise in the second period ranking, like e.g. "Infrared, submillimeter wave, microwave and radiowave instruments, equipment and techniques", "Sensors (chemical, optical, electrical, movement, gas, etc)", "Optical elements, devices and systems", "Material sciences: Methods of deposition of films and coatings ; film growth and epitaxy" (see in table 2, the ranks 1 to 4, 6 to 9, and 12 to 15), and 4 fields lose places in the ranking of the second period (see in table 3 the fields written in italic): "Physico-chemistry of polymers: Organic polymers", "Polymer industry: Technology of polymers", "Electronics: Materials", "Electric, optical and optoelectronic circuits".

Four fields appear in the top 20 field list in the second period: "Nanoscale materials and structures: Fabrication and characterization", "Structure and nonelectronic properties of surfaces, interfaces and thin films", "Natural energy", "Materials science" and, conversely, four fields disappear from the top 20 field list: "Industrial chemicals", "Electrical engineering. Electrical power engineering", "General and physical chemistry", "Computer science: Software".

---

[1] 2010 can not be complete: the data, recorded on June 2010, contains around a half of the expected production of one entire year

ACOUSTICS.
ANALYTICAL CHEMISTRY.
ATOMIC AND MOLECULAR PHYSICS.
BUILDINGS. PUBLIC WORKS.
CHEMICAL ENGINEERING.
CHEMICAL INDUSTRY AND CHEMICALS.
COMPUTER SCIENCE. CONTROL THEORY. SYSTEMS.
CONDENSED MATTER: ELECTRONIC STRUCTURE, ELECTRICAL,
CONDENSED MATTER: STRUCTURE, MECHANICAL AND THERMAL
CROSS-DISCIPLINARY PHYSICS: MATERIALS SCIENCE, RHEOLOGY.
EARTH, OCEAN, SPACE.
ELECTRICAL ENGINEERING. ELECTRICAL POWER ENGINEERING.
ELECTROMAGNETISM, ELECTRON, AND ION OPTICS.
ELECTRONICS.
ENERGY.
FLUID DYNAMICS.
FUNDAMENTAL AND APPLIED BIOLOGICAL SCIENCES.
GENERAL AND PHYSICAL CHEMISTRY.
GROUND, AIR AND SEA TRANSPORTATION, MARINE
HEAT TRANSFER.
INFORMATION SCIENCE. DOCUMENTATION.
INFORMATION, SIGNAL AND COMMUNICATION THEORY.
INORGANIC CHEMISTRY AND ORIGINS OF LIFE.
INSTRUMENTS, APPARATUS, COMPONENTS AND TECHNIQUES
MATHEMATICS.
MECHANICAL ENGINEERING. MACHINE DESIGN.
MEDICAL SCIENCES.
METALS. METALLURGY.
NUCLEAR PHYSICS.
OPERATIONAL RESEARCH. MANAGEMENT SCIENCE.
OPTICS.
ORGANIC CHEMISTRY.
PHYSICOCHEMISTRY OF POLYMERS.
PHYSICS OF ELEMENTARY PARTICLES AND FIELDS.
PHYSICS OF GASES, PLASMAS AND ELECTRIC DISCHARGES.
POLLUTION.
POLYMER INDUSTRY, PAINTS, WOOD.
SOLID MECHANICS.
TELECOMMUNICATIONS.

Table 1. General scientific fields present in the corpus "Optoelectronic devices"

Rank	PASCAL classification scheme headings
1	Semiconductor electronics. Microelectronics. Optoelectronics. Solid state devices
2	Physico-chemistry of organic polymers
3	Polymer industry: Technology of polymers
4	Infrared, submillimeter wave, microwave and radiowave instruments, equipment and techniques
5	Electronics: Materials
6	Condensed matter: Optical properties and spectroscopy
7	Industrial chemicals
8	Electric, optical and optoelectronic circuits
9	Lasers
10	Electrical engineering. Electrical power engineering
11	Optical elements, devices and systems
12	Sensors (chemical, optical, electrical, movement, gas, etc)
13	Structure of solids and liquids; crystallography
14	Material sciences: Methods of deposition of films and coatings; film growth and epitaxy
15	General and physical chemistry
16	Condensed matter: Electronic structure and electrical properties of surfaces, interfaces and thin films
17	Computer science: Software
18	Optical instruments, equipments and techniques
19	Optical materials
20	Optical sources and standards

Table 2. The most important fields in the first period (2000-2004) expressed by means of the PASCAL classification scheme headings

At first sight, these evolutions suggest a decline in P2 of the number of records dealing with the characterization of materials, for the benefit of documents dedicated to the study of devices, as well as the emergence of new forms of materials as thin films and nanomaterials. More than 200 journals contributed to form each of the two studied corpus. In the first period, 9,666 authors working in institutions located in 69 different countries produce the 2,950 obtained records. In the second period, the number of authors is almost multiplied by 2 and the number of affiliation countries also rise to 81. In table 4, we show the top 10 affiliation countries for both periods. If the countries present in both lists are the same, their rankings are quite different. The USA remains in the first place with an almost unchanged rate, but we can see the remarkable ascent of China that doubles its affiliation rate coming from the sixth to the second place. With the exception of Canada, all other countries see their affiliation rate decreasing.

Rank	PASCAL classification scheme headings
1	Semiconductor electronics. Microelectronics. Optoelectronics. Solid state devices
2	Condensed matter: Optical properties and spectroscopy
3	Infrared, submillimeter wave, microwave and radiowave instruments, equipment and techniques
4	Sensors (chemical, optical, electrical, movement, gas, etc)
5	*Physico-chemistry of organic polymers*
6	Optical elements, devices and systems
7	Material sciences: Methods of deposition of films and coatings; film growth and epitaxy
8	Optical materials
9	Lasers
10	Nanoscale materials and structures: Fabrication and characterization
11	*Electric, optical and optoelectronic circuits*
12	Condensed matter: Electronic structure and electrical properties of surfaces, interfaces and thin films
13	Optical sources and standards
14	Structure and nonelectronic properties of surfaces, interfaces and thin films
15	Optical instruments, equipments and techniques
16	Structure of solids and liquids; crystallography
17	*Polymer industry: Technology of polymers*
18	Natural energy
19	Materials science
20	*Electronics: Materials*

Table 3. The most important fields in the second period (2005-2010) expressed by means of the PASCAL classification scheme headings

First period			Second period		
Rank	Country	% of affiliations	Rank	Country	% of affiliations
1	USA	28	1	USA	29
2	Japan	13	2	China	18
3	UK	10	3	Japan	10
4	Germany	10	4	Taïwan	9
5	South Korea	10	5	South Korea	8
6	China	9	6	Germany	7
7	France	7	7	UK	6
8	Italy	6	8	France	6
9	Taïwan	6	9	Canada	4
10	Canada	3	10	Italy	3
… 69			… 81		

Table 4. The top 10 countries in terms of affiliation rates in the first and second periods

This literature is of English expression in more than 99 % of documents and was published in 14 and 11 countries, respectively, in the first and second period. In table 5, we give the top 5 publishing countries for both periods. The USA strengthens its first place, the Netherlands rate grows up and the rates of Switzerland, United Kingdom and Germany go down.

First period			Second period		
Rank	Country	% of publications	Rank	Country	% of publications
1	USA	41	1	USA	58
2	Switzerland	16	2	Netherlands	15
3	UK	14	3	UK	12
4	Germany	12	4	Switzerland	9
5	Netherlands	11	5	Germany	4
… 14			… 11		

Table 5. The top 5 countries in terms of publishing rates in the first and second periods

3. Methodology

3.1 Diffusion model

New scientific and technological developments are presented and documented in scientific publications. These new ideas are described using specific words and expressions and they sometimes lead to the creation by scientists of new terminological representations. In a bibliographic record, the quintessence of a publication is represented by means of keywords, the investigation of which is meaningful for tracking the developments of new technologies or new ideas in science, especially if these keywords are given by the authors.

Our model is based on the consideration that technologies find 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, the new terminology is used in few publications and appears therefore in few bibliographic records. New terms are created or applied coming from other disciplines with sometimes a new meaning. If the findings are of interest for other research groups, they pick them up and the number of publications, and of records, including these terms grows. In this second stage the findings are primarily interesting for researchers in the same discipline and the growth is essentially restricted to this area. If the findings are also considered as fundamental for the use in another discipline, the related terms diffuse over the scientific community and they occur in publications of many disciplines. By investigating the terminology of a suitable corpus of bibliographic records at a certain time span, we can divide the terminology into three stages, which expresses an invention or new findings, growth in a discipline and diffusion in the scientific community:

- 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 where they form an exotic set of unusual terms.

- Stage 2: Already established specific terms in the considered technology (or discipline, etc.) occur in this stage 2. We refer to these terms as established terms.

- 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 three-stage model the "Diffusion Model". Therefore, the diffusion model identifies, with the help of simple criteria, three categories of terms: unusual terms in a considered field, established terms in a considered field, and cross-section terms regarding to a considered field. The categorisation of a term is based on the following indicators and their related criteria:

	STAGE 1	STAGE 2	STAGE 3
Consider a specific term for the following indicators	**Unusual terms in a considered field**	**Established terms in a considered field**	**Cross-section terms regarding a considered field**
Total frequency	low	high	high
Number of fields where the term occurs (diffusion of a term)		in less than 1/3 of all fields	in more than 1/3 of all fields
Ratio of the frequency in the considered field to total frequency (which means that a term focuses in a specific technology – Field Focus)	more than 2/3	more than 2/3	less than 0.9
Age of the term	young	old	old

Table 6. Criteria to dispatch the terms of a considered field into the three stages

These criteria are linked to the corpus characteristics and, in the framework of a diachronic analysis as ours, their numerical value can differ within each period (see table 7 and 9).

The "total frequency" of a term represents the number of different bibliographic records in which it occurs at least once. Each bibliographic record is associated to at least one scientific field. The more specific a publication is, the fewer the scientific fields associated with its bibliographic record. The fewer scientific fields a term is associated with, the more the term is specific.

The "field focus" of a term is an indicator of its specificity and it lies between 0 and 1. The higher the field focus of a term, the more the term seems to focus on a specific field or the less it is spread to other fields.

The "age of the term" is defined in relation with the year when it occurs in a bibliographic record for the first time in the considered field for a given period. It is considered old if it appears during the first years of the period and young otherwise.

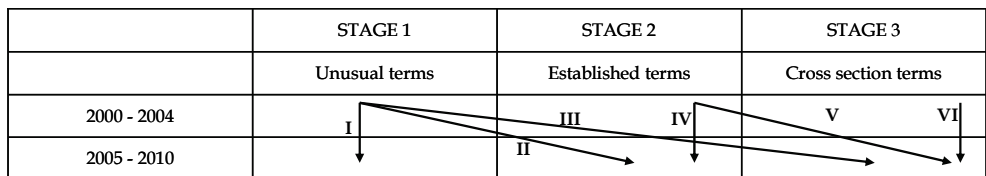|	STAGE 1	STAGE 2	STAGE 3
	Unusual terms	Established terms	Cross section terms
2000 - 2004	I	III IV	V VI
2005 - 2010		II	

Fig. 1. Taxonomy of the migration pathways of terms through different stages of the diffusion model in two time periods

In order to dispatch the terminology between the three stages, each four indicators given in table 6 has to be determined for each term. The dynamics of emerging technologies can be described by pathways of terms 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 terms that follow the mentioned pathways, see Besagni et al. (2009).

Following the diffusion approach it can be assumed that terms migrate from earlier stages in the first period to later stages in the second period like the following:

- terms from stage 1 in the first period stay in this stage or migrate in the second period to stage 2 or stage 3
- terms from stage 2 in the first period stay in stage 2 or migrate to stage 3
- terms from stage 3 stay in stage 3

We define pathways as follows:

- Pathway I: the same terms remain in stage one in both periods
- Pathway II: terms move from stage 1 in first period to stage 2 in second period
- Pathway III: terms move from stage 1 in first period to stage 3 in second period
- Pathway IV: the same terms remain in stage two in both periods
- Pathway V: terms move from stage 2 in first period to stage 3 in second period
- Pathway VI: the same terms remain in stage three in both periods

For the visualization and analysis we used the software BibTechMon™, a bibliometric monitoring tool (Kopcsa & Schiebel 1998). The core of the software is an algorithm to position related objects (in our case, terms) in a two dimensional space. The relation between terms is defined by their pairwise co-occurrence in bibliographic records. 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. Thus, a term is positioned with respect to all other ones. If a term in the centre of the network or of a subnetwork has a rather strong relationship to many other nodes in the network graph, that is called a high centrality.

In our case, established terms or cross-section terms are centred. New terms and unusual terms are positioned in the periphery. A bunch of terms that are cited often together crossing over many publications forms up a subnetwork.

3.2 Cluster analysis

The diachronic cluster analysis is based on the study of the evolution of cluster maps and contents obtained by a clustering method and considering data sets related to two or more successive time periods. In this work, we deal with two time periods and we apply a clustering algorithm on the corpus of each period, in which documents are represented by the keywords existing in the bibliographic records. The clustering tool we use (Lelu & François 1992; Lelu 1993) applies a non-hierarchical, non-supervised clustering algorithm, the axial K-means method that allows producing clusters presenting the following particular characteristics:

- they can overlap because the clustering method allows a document or a keyword to belong to more than one cluster,
- documents and keywords as the constituting elements of a cluster, are ranked by decreasing similarity with the cluster ideal type.

The clustering is operated by the software-tool Stanalyst (Polanco et al. 2001) that produces a graphical representation of the clustering results under the form of a network of clusters.

Each cluster is a homogeneous subset of the information contained in the analyzed corpus, corresponding to a topic. After the clustering, we represent graphically its results employing two methods: a principal component analysis and a connected component analysis. The first one produces a global map of the topics. The topics that are very far apart represent clusters very different with respect to the keywords defining them.

The connected component analysis is then employed to show the relationships that exist between the clusters. It is a method based on the graph theory and defining the connected components representing the relative proximity between clusters. The distance between two clusters is defined as the cosine of the angle between the two axes representing the clusters. A lower and an upper threshold are defined as respectively the minimum and the maximum value of the cosine between the clusters. Nine intermediate levels are defined in that interval. We consider that, at a given level, two clusters are connected if the value of the cosine associated to them is within the range of values of this level.

Finally, in order to analyse the evolution of the cluster vocabulary between the two considered periods, we build a comparison matrix (Roche et al. 2008) 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. These detected phenomena need to be validated by experts from the technological domain.

The cluster behaviours are calculated in the following way. Let M be the comparison matrix:

$$M = (m_{i,j}) \text{ with :}$$

$i = 1, \dots, n_{P2}$ where n_{P2} = number of clusters obtained in the second period

$j = 1, \dots, n_{P1}$ where n_{P1} = number of clusters obtained in the first period

m_{ij} = percent of keywords belonging to cluster i and existing in cluster j

and, for each line i, the marginal value is:

$$. \ m_{i.} = \sum_{j=1}^{j=n_{P1}} m_{ij} \ .$$

The analysis of the matrix M allows us to build the following hypotheses:

- Let (i,k) be a pair of homonymous clusters. The cluster i is supposed to be stable if:

$$m_{ij} < \frac{m_{ik}}{2} ; \forall j \neq k \text{ and } j = 1, n_{P1} \tag{1}$$

- Let (i,k) be a pair of homonymous clusters. The cluster i is supposed to be unstable if there exists at least one j respecting the condition:

$$m_{ij} > \frac{m_{ik}}{2} ; \forall j \neq k \text{ and } j = 1, n_{P1} \tag{2}$$

- Clusters having new titles (titles not existing in the first-period cluster list) and obeying the condition (1) are supposed to be stable;

- Clusters with the lowest marginal values have less inheritance from first-period clusters and we suppose they constitute new terminological notions;
- Clusters with the highest marginal values are supposed to aggregate a great number of notions already existing in the first-period clusters.

At this point, scientific expertise is needed. We ask scientific experts to validate our hypotheses regarding the analysis of the two cluster maps, taking into account the cluster contents.

4. Results

4.1 Diffusion model

The diffusion model has been applied to the 8,169 bibliographic records delivered by the PASCAL database query, split into two corpora according to their publication year: 2,950 references for P1 (2000 to 2004) and 5,219 references for P2 (2005 to 2010). In these bibliographic records:

- the technological fields are the PASCAL classification codes, reduced to the classification scheme headings,
- the terms are the PASCAL keywords,
- a term is considered old if it appears during the first 3 years of a period.

For P1, the terms have been dispatched into the three stages according to the following criteria:

	STAGE 1	STAGE 2	STAGE 3
Consider a specific term for the following indicators	**Unusual terms in a considered field**	**Established terms in a considered field**	**Cross-section terms regarding a considered field**
Total frequency	between 8 and 20	> 40	> 30
Number of fields where the term occurs (diffusion of a term)		< 17	≥ 17
Ratio of the frequency in the considered field to total frequency - Field Focus	> 0.7	> 0.7	< 0.9
Age of the term	young	old	old
Year of first occurrence	2003-2004	2000-2002	2000-2002

Table 7. Filter for the corpus of the first period (2000 – 2004) for 2,950 records

This filter does not allow for an overlap of common terms. Each term is assigned to only one of these three stages.

The criteria are determined as follows:

- Stage 1: The total frequency is defined higher than 7 in order to reject badly spelt terms. To exclude established terms, the total frequency upper bound is fixed to 20. As we obtain 2,590 records for the first period, a frequency of 20 means that such a term occurs in less than 0.8% of the documents of P1. Furthermore a term must be relevant and not used by chance. The value of field focus should then be high, at least 0.7, so that a focus on a field is ensured.
- Stage 2: Established terms must have a higher total frequency than in stage 1. Therefore we fixed the lower bound to 40 in P1 (which means a term frequency of 1.5%). An

established term should focus on a specific field. Therefore it has to occur in less than 17 fields, which corresponds to a third of the total number of fields of P1. The spread to other fields (the field focus) should be high, here 0.7.

- Stage 3: A cross-section term has a high total frequency : it is used in more than 40 records in P1. Furthermore, it must be used in a high number of fields, in P1 more than 17. The field focus should ensure a distribution to many fields and must not exceed 0.9.

This filtering results in the following table, presenting the top 10 terms of each stage:

Unusual terms	Established terms	Cross-section terms
CMOS image sensors	Binary compound	Photoluminescence
Thermal imaging	Semiconductor materials	Ternary compound
Sensor materials	Quantum well	Thin film
Imager	Gallium nitride	Conjugated copolymer
Chip	III-V semiconductors	Quantum yield
Test	Optical absorption	Phenylenevinylene derivative polymer
Spatial resolution	Molecular beam epitaxy	Measurement sensor
Digital image	Gallium arsenides	Phenylene derivative copolymer
Organometallic compound	Gallium Nitrides	Crystal growth from vapors
Diode array	Injection laser	Silicon

Table 8. Top 10 terms for the first period (2000 – 2004)

The P2 terms have been dispatched into three stages according to the following criteria:

	STAGE 1	STAGE 2	STAGE 3
Consider a specific term for the following indicators	Unusual terms in a considered field	Established terms in a considered field	Cross-section terms regarding a considered field
Total frequency	between 8 and 40	> 80	> 50
Number of fields where the term occurs (diffusion of a term)		< 16	≥ 16
Ratio of the frequency in the considered field to total frequency – Field Focus	> 0.7	> 0.7	< 0.9
Age of the term	young	old	old
Year of first occurrence	2008- 2010	2005-2007	2005-2007

Table 9. Filter for the corpus of the second period (2005 – 2010) for 5,219 records

The criteria are similar to the first period (see below Table 7) but adapted to the number of records of 5,219 records of this second period. Therefore the numbers of the total frequency differs from them in the first period.

This filtering results in table 10, presenting the top 10 terms of each stage.

Unusual terms	Established terms	Cross-section terms
Aluminium Nitrides	Voltage current curve	Electroluminescence
Electron transport	Binary compound	Photoluminescence
Oxine	Electrochemical properties	Thin film
Amplifier	Optical material	Quantum yield
Biomedical imaging	Brightness	III-V semiconductors
Computer software	Focal plane arrays	Nanostructure
Electronic conductivity	Luminescent material	Measurement sensor
Gallium Phosphides	Quantum well	Optical properties
Daylight	Conjugated polymer	Silicon
Holes	Thermal stability	Conjugated copolymer

Table 10. Top 10 terms for the second period (2005 – 2010)

4.1.1 Pathways

In that study dealing with the identification of emergent research issues, it appears that the Pathways II, III and V carry the bulk of the relevant information highlighting the potential scientific evolutions, or even breakthroughs. That is why we focus on these three pathways.

4.1.1.1 Pathway II

Terms which follow the Pathway II progress from "unusual term" in the first period to "established term" in the second period.

	STAGE 1	STAGE 2	STAGE 3
	Unusual terms	Established terms	Cross section terms
2000 - 2004		II	
2005 - 2010			

Fig. 2. Migration Pathway II

The terms which follow Pathway II are:
- *Sensor materials*
- *Thermal imaging*

These two keywords play a role in several articles as:
- Advanced pixel design for infrared 3D LADAR imaging
- Advanced infrared detectors for multimode active and passive imaging applications
- Dual-Band MW/LW IRFPAs made from HgCdTe grown by MOVPE
- Comparison of QWIP to HgCdTe detectors for gas imaging

4.1.1.2 Pathway V

Terms which follow the Pathway V progress from the "established stage" in P1 to the "cross-section stage" in P2.

These terms are:
- *Doping*
- *Gallium arsenides*
- *III-V semiconductors*
- *Injection laser*
- *Luminescence*
- *Quantum dot*

For example, the keywords "*Injection laser*", "*Luminescence*", or" *Quantum dot*" are used to index publications like:

- Carrier dynamics in active-region materials for diode laser applications.

The keywords "*Gallium arsenides*", "*Injection laser*" or "*Quantum dot*" are used in:

- Superluminescent Diodes versus Broadband Laser Diodes.

	STAGE 1	STAGE 2	STAGE 3
	Unusual terms	Established terms	Cross section terms
2000 - 2004			V
2005 - 2010			

Fig. 3. Migration Pathway V

4.1.1.3 Pathway III

Terms which follow the Pathway III progress from the "unusual stage" in P1 to the "cross-section stage" in P2.

	STAGE 1	STAGE 2	STAGE 3
	Unusual terms	Established terms	Cross section terms
2000 - 2004		III	
2005 - 2010			

Fig. 4. Migration Pathway III

Pathway III results in the following terms:
- *Imager*
- *Nanoparticle*

The keyword "*Imager*" is used in publications as:
- Gated Diode Design to Mitigate Radiation Damage in X-Ray Imagers
- A Day and Night Vision MOS Imager With Robust Photonic-Crystal-Based RGB-and-IR.

The keyword "*Nanoparticle*" indexes publications as:
- Identification of unamplified genomic DNA sequences using gold nanoparticle probes and a novel thin film photodetector
- Direct gravure printing of indium tin oxide nanoparticle patterns on polymer foils
- Photoluminescence from ZnO Nanoparticles Embedded in an Amorphous Matrix
- Use of surface plasmon-coupled emission for enhancing light transmission through Top-Emitting Organic Light Emitting Diodes

4.1.2 BibTechMon™ visualization

For the calculation and the visualization of the term networks we used the software BibTechMon™. The algorithm in BibTechMon™ is based on a co-object analysis, which can be briefly described as follows. An object (in our case, a term) occurs with a certain frequency, defined by the number of bibliographic records in which the considered term occurs at least once. It is counted once per record. When a term occurs with another term in the same bibliographic record, we call that a co-occurrence. The co-occurrence strength is given by the number of co-occurrences pairwise. These co-occurrences are normalized via the so called Jaccard index. After that, the terms find their position inside the network by the application of a spring model, where the Jaccard indices are the measures for the interactions. Thus, a term is positioned with respect to all other terms and to their Jaccard indices.

The network consists of points which represent the terms and lines which show the intensity of the individual relation between two terms. The size of the points represents the total frequency of the term occurring in the corpus. The relative position of the objects compared to each other is important: objects which are positioned close to each other form subtopics and clusters because they occur in similar contexts.

Figure 5 shows the location of the terms of Pathway V in the network of the second period. In this figure, cross-section terms can be found in the subnetwork at the centre. They have connections to many other terms.
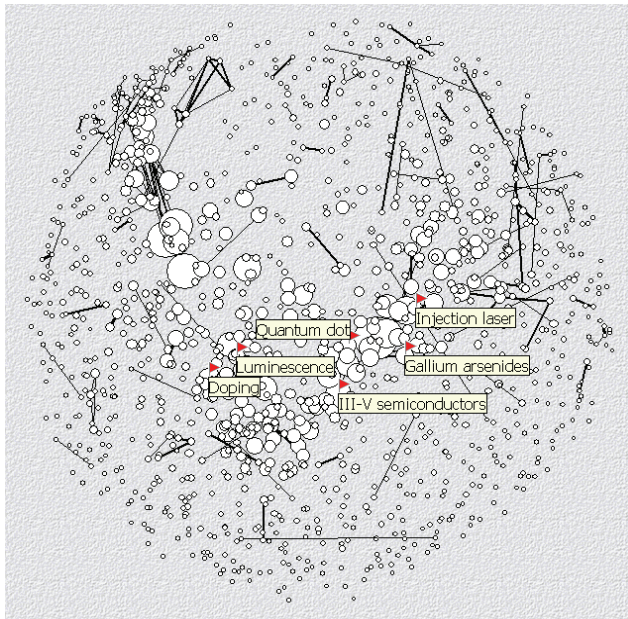

Fig. 5. Pathway V terms in the network of the second period

The study of Pathway III is even more interesting. Figures 6 and 7 show respectively the terms of Pathway III ("*Nanoparticle*", "*Imager*") in the network of the first period and the second period.
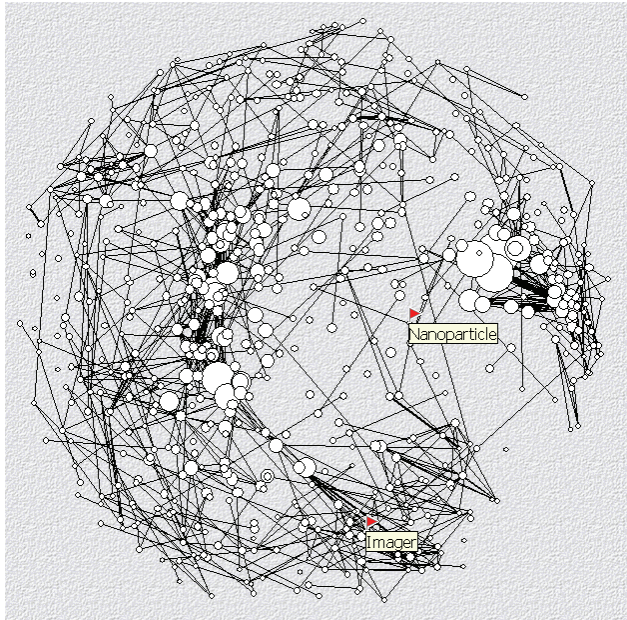
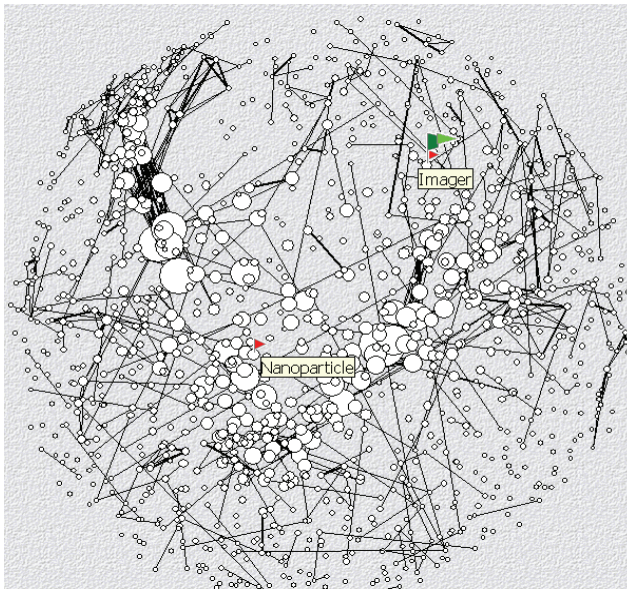Fig. 6. Terms of Pathway III in the first period

Fig. 7. Terms of Pathway III in the second period

In the first period, the terms "*Nanoparticle*" and "*Imager*" are unusual. Indeed, their position is marginal with respect to the strong subnetworks of the P1 graph (figure 6). Conversely,

these terms play a more important role in the second period: they are located in the core of important and dense networks of the P2 graph (figure 7). Furthermore, as figure 8 shows, these two terms are connected to many other terms in the network of the second period.
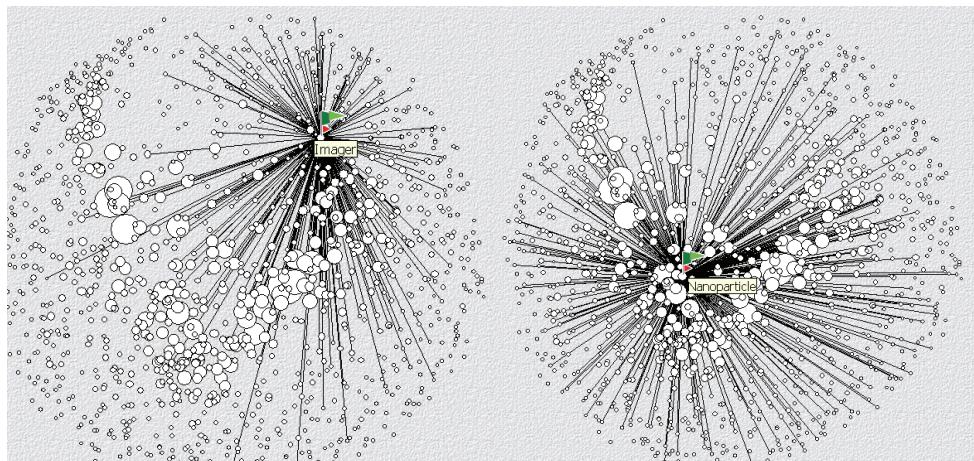

Fig. 8. Networks of "*Imager*" and "*Nanoparticle*" in the second period

The visualization obtained with BibTechMon™ thus supports the results of the diffusion model according to Pathway III and Pathway V that is the strong networks built by imaging techniques and nanotechnology in the optoelectronic technological field for the second period.

4.2 Cluster analysis

The 8,169 records delivered by the PASCAL database query have been split into two corpora according to their publication year: 2000 to 2004 (P1) and 2005 to 2010 (P2). In the first period, we obtained 2,950 documents and 6,735 different keywords among which 5,257 (78%) appear less than 4 times. The second period had 5,219 articles and 6,378 keywords among which 3,976 (62%) occur less than 4 times.

The clustering step was preceded by an expert revision of the keyword index of each period corpus in order to detect the presence of indexing artefacts, to homogenize the variational forms of common concepts and to eliminate keywords that are not discriminating in the context of our analysis. For instance, all the keywords employed in the query have been discarded. On the other hand, the expert decided to take into account only the keywords occurring at least 4 times in the two corpora. After this terminology validation step, we obtained for P1 and P2 two indexes formed, respectively, by 1,472 and 2,391 keywords that were the input data of the clustering step.

After that, we study the evolution of the clusters between P1 and P2 with the help of a comparison matrix before its validation by the experts.

4.2.1 Clustering step

We applied a clustering algorithm on each one of our two corpora in which the documents are represented by the keywords existing in the bibliographic records. In the table 11, we

present, for the two periods, the name of the obtained clusters and the number of theirs keywords and documents. The homonymous clusters are highlighted.

	Nb keywords	Nb records	Second period	Nb keywords	Nb records	First period
Cluster 1	27	363	Organic light-emitting devices	90	124	Optical telecommunication
Cluster 2	21	725	Infrared detector	22	408	Infrared detector
Cluster 3	41	881	Electroluminescence	64	45	Image processing
Cluster 4	46	192	Single photon	142	162	Chemical synthesis
Cluster 5	80	125	Optical method	61	150	Semiconductor quantum wells
Cluster 6	67	136	Optical sensor	34	338	Semiconductor materials
Cluster 7	57	264	Semiconductor laser	14	454	Conjugated polymer
Cluster 8	73	483	Thin film	91	59	Printed circuit board
Cluster 9	144	134	Nonimaging optics	25	242	III-V semiconductors
Cluster 10	70	209	Lighting fitting	21	637	Electroluminescence
Cluster 11	42	448	III-V semiconductors	57	201	Thin film
Cluster 12	38	542	Conjugated copolymer	47	68	CCD image sensors
Cluster 13	38	379	Measurement sensor	31	458	Conjugated copolymer
Cluster 14	81	174	Integrated optics	96	66	Computer simulation
Cluster 15	94	172	Photonic crystal	94	89	Photons
Cluster 16	70	138	Radiation sources	20	583	Photoluminescence
Cluster 17	53	406	Nanostructure	45	156	Semiconductor laser
Cluster 18	55	249	Europium addition	58	148	Measurement sensor
Cluster 19	97	206	Photodiode	60	70	LED lamps
Cluster 20	22	1047	Photodetector	69	110	Optical sensor

Table 11. Clusters of the two periods: P1 (2000-2004) and P2 (2005-2010)

The clustering algorithm produced 20 clusters involving 2,938 documents and 823 keywords in P1, and 20 clusters relating 5,214 documents and 937 keywords in P2. Let us notice that the important reduction, in both periods, of the number of keywords associated to the clusters, around 85% 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, despite these huge reductions of terminology, we always had almost 100% of the initial information corpora in the clusters. The figures 9 and 10 show the cluster maps obtained for, respectively, P1 and P2.
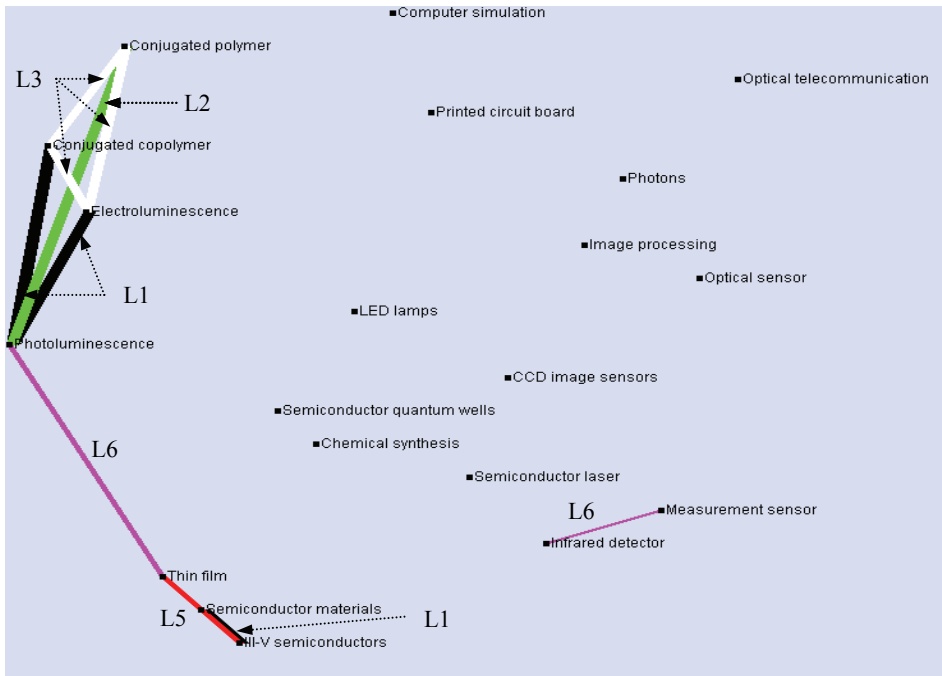
Fig. 9. Cluster map of P1 (2000-2004), with indication of the strength of the connections
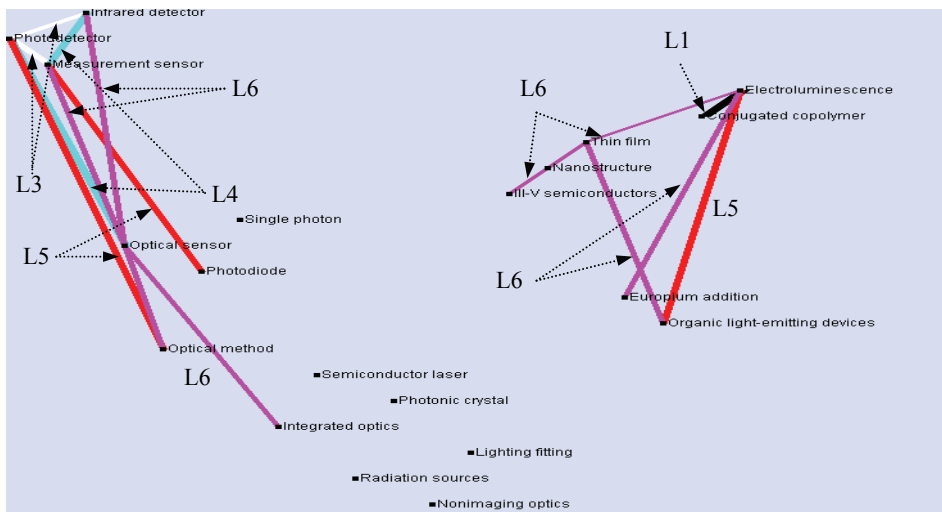

Fig. 10. Cluster map of P2 (2005-2010), with indication of the strength of the connections

In both maps produced by Stanalyst, each dot corresponds to a cluster and each line gives the connexion level between pairs of clusters. Note that we considered here only the 6

highest levels (out of the 10 defined by the connected component analysis step) corresponding to the strongest connexions between clusters.

In Stanalyst, the level of connexions is code-coloured. Since these colours do not come out in print, we mentioned in figure 9 and 10 the level next to the connexion with a symbols going from L1 (strongest connection) to L8 (weakest connection).

4.2.2 Comparison matrix analysis

The main purpose of the diachronic cluster analysis is to determine which topics of the second period find their roots in the first one and which new topics emerge in the second period. In order to analyze the evolution of the cluster vocabulary between the two considered periods, we built 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.

The comparison matrix is showed in table 12. For instance, if we consider the row 17 of the matrix, corresponding to the cluster "Nanostructure" of P2, we can see that 26% (see element highlighted) of its keywords come from the cluster "Chemical synthesis" (see column 4) of P1. Furthermore, the cumulated values of the set of inheritances of each cluster of P2, in terms of keywords already present in the clusters of P1, are given in the marginal column.

P2/P1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	3	0	3	7	0	0	0	0	0	7	22	0	0	0	3	7	7	0	7	0	*66*
2	0	42	4	0	0	4	0	0	0	0	0	4	0	0	0	0	0	9	0	0	*63*
3	0	0	0	4	2	0	24	0	0	21	4	0	26	0	0	34	0	0	7	0	*122*
4	2	6	2	2	0	4	0	2	6	0	0	2	0	2	4	0	4	0	0	0	*36*
5	3	1	8	2	0	0	0	1	0	0	0	3	0	3	2	0	2	7	5	12	*49*
6	4	1	8	5	0	0	0	0	0	0	0	5	0	2	2	0	1	8	1	11	*48*
7	8	5	1	0	3	5	0	0	3	0	0	1	0	3	7	0	26	0	1	3	*66*
8	0	0	0	12	4	5	4	0	5	4	31	0	4	0	5	5	0	1	1	0	*81*
9	2	0	11	0	0	0	0	3	0	0	0	6	0	5	1	0	0	4	7	3	*42*
10	1	0	5	0	0	0	0	2	0	1	0	1	2	4	0	1	2	4	51	4	*78*
11	0	7	2	7	11	33	4	0	40	2	14	0	4	0	0	4	7	0	0	0	*135*
12	0	0	0	0	2	0	13	0	0	5	2	0	73	0	0	15	0	0	5	0	*115*
13	2	10	2	0	0	0	0	2	0	0	5	5	0	2	0	0	0	18	0	7	*53*
14	12	0	4	0	0	1	0	6	0	0	0	1	0	7	9	0	11	3	0	12	*66*
15	8	1	4	1	1	0	0	2	1	0	4	1	0	7	15	0	7	0	2	3	*57*
16	5	0	8	5	4	0	0	7	0	0	0	0	0	5	8	0	7	1	7	7	*64*
17	1	1	3	26	1	1	5	0	1	1	9	1	5	0	5	5	1	0	0	1	*67*
18	1	0	1	10	0	7	1	0	0	1	12	3	1	1	3	3	1	0	9	0	*54*
19	4	1	4	3	0	1	0	3	1	0	1	9	0	5	3	0	1	12	4	5	*57*
20	4	13	0	0	4	9	0	0	4	0	4	13	0	9	13	0	9	13	0	4	*99*

Table 12. Comparison matrix of the keywords of the two periods: P1 (2000-2004) and P2 (2005-2010)

The analysis of the comparison matrix allows us to singularize in the second period (see table 13):

- eight homonymous clusters, but only four of them seem to be stable
- three clusters with new titles but presenting characteristics of stability
- four clusters with a low (less than 50%) inheritance global rate from the clusters of P1 - *Are they new clusters?*
- three clusters with a strong (more than 100%) inheritance global rate from different clusters of P1 – *Do they convey "old", already known information?*

Type of singularization	Cluster name
Homonymy with P1, apparent stability (larger keywords inheritance coming from the homonymous cluster of P1)	Infrared detector, Semiconductor laser, Thin film, Conjugated copolymer
Homonymy with P1, less apparent stability (larger keywords inheritance coming from other clusters of P1 than the homonymous one)	Electroluminescence, Optical sensor, III-V semiconductors, Measurement sensor
New title with stability characteristics	Organic light-emitting devices, Lighting fitting, Nanostructure
Low inheritance global rate	Single photon, Nonimaging optics, Optical sensor, Optical method
High inheritance global rate	III-V semiconductors, Conjugated copolymer, Electroluminescence

Table 13. Clusters singularized in second period

4.2.3 Experts validation

Starting from these quantitative data the expert worked on the validation of our hypotheses by analysing the contents of the clusters and their relative position in the cluster network drew in the map of each period. Special attention has been paid on the detection of brutal changes of cluster relative position.

The expert's analysis confirms the stability of two homonymous clusters between P1 and P2: "Infrared detector" and "Semiconductor laser". The homonymous clusters "Thin film" and "Conjugated copolymer" are rather stable between P1 and P2, but the former shows a higher relative weight of organic towards inorganic materials in P2, and the latter a higher relative weight of devices towards characterization of the materials.

Despite their less apparent stability, due to a lower inheritance from P1 to P2, the homonymous clusters "Measurement sensor" and "III-V semiconductors" are stable between P1 and P2. The other clusters of P1 from which they inherit deal with similar topics.

The P2 cluster "Optical sensor" does not inherit predominantly from its P1 homonymous. Meanwhile, its stronger keywords inheritance comes from topically close clusters as "Image processing" and "Measurement sensors". This cluster is rather stable between P1 and P2, but it highlights the emergence in P2 of some applications related to biomedical imaging and security. The take off of these new applications of sensors can explain the relatively low inheritance level of the P2 clusters "Optical sensor" and "Optical method".

As the former one, the P2 cluster "Electroluminescence" does not inherit predominantly from its P1 homonymous. Indeed, three P1 clusters hand more keywords down to it than its homonymous: "Photoluminescence", "Conjugated polymers" and "Conjugated copolymers". This fact shows the strong inheritance relation between the studies of P1 related to synthesis of conjugated (co)polymers and characterization of their luminescent properties, and the studies of P2 related to the potential applications of these materials for the design of electroluminescent devices.

The strong inheritance of the P2 cluster "Lighting fitting" from the P1 cluster "LED lamps" shows the importance of the topic dealing with the use of electroluminescent devices as light sources through the two periods of time. In the second period, the weight of this topic is reinforced by the presence of the cluster "Nonimaging optics" and by a growing number of bibliographic records.

The P2 cluster "Organic light-emitting devices" inherits most of its keywords from the P1 cluster "Thin films". This relation shows the temporal evolution of the field, from the study of the potential use of thin films and heterostructures in optoelectronic devices in P1, to the fabrication and characterization of these devices in P2. The stronger implication of organic materials in P2, highlighted by the name of the P2 cluster, confirms the tendency formerly observed when analyzing the stability of the homonymous class "Thin films" between P1 and P2.

The P2 cluster "Nanostructure" inherits most of its keywords from the P1 cluster "Chemical synthesis", which deals essentially with the synthesis of nanomaterials (nanowires, nanocomposites, nanoboxes, nanospheres…). This inheritance, together with the strong relative growth of the number of documents between the two clusters, shows the growing importance of nanomaterials in the field of optoelectronic devices in P2.

Amongst the clusters of P2 showing a low marginal value, "Single photon" gathers bibliographic records dealings with the application of avalanche photodiodes to fields like quantum information, quantum communication or quantum cryptography. This low marginal value is thus closely related to the emergence of a new scientific field.

The analysis of the cluster maps obtained for the two periods completes and confirms these observations.

In the cluster map of the first period (see figure 11), the four cluster "Conjugated polymers", "Conjugated copolymers", "Electroluminescence" and "Photoluminescence" form a very strong network (N1), This result highlights the importance of organic semiconductors in the field of optoelectronic devices and confirms the conclusion of a former study, in which we analyzed the evolution of the same scientific field over an older time period (Schiebel et al. 2009). This network is still present in the second period, but it has included the cluster "Organic light-emitting device", illustrating the transition from materials to devices. Furthermore, the proximity of the cluster "Nanostructure" shows the growing implication of such materials in optoelectronic devices in P2.

The topic related to imaging devices is already present in the first period (see figure 11), but the corresponding clusters ("Measurement sensor", "Infrared detector", "Optical sensor", "Image processing") do not show strong relations. On the contrary, they form a much strong network (N2) in the second period (see figure 12). Though this topic is not new, the comparison of the two cluster maps give evidence of its growing importance in the second period. Close to this network, the appearance in the second period of the cluster "Single photon" illustrates the emergence of quantum information processing in the field.
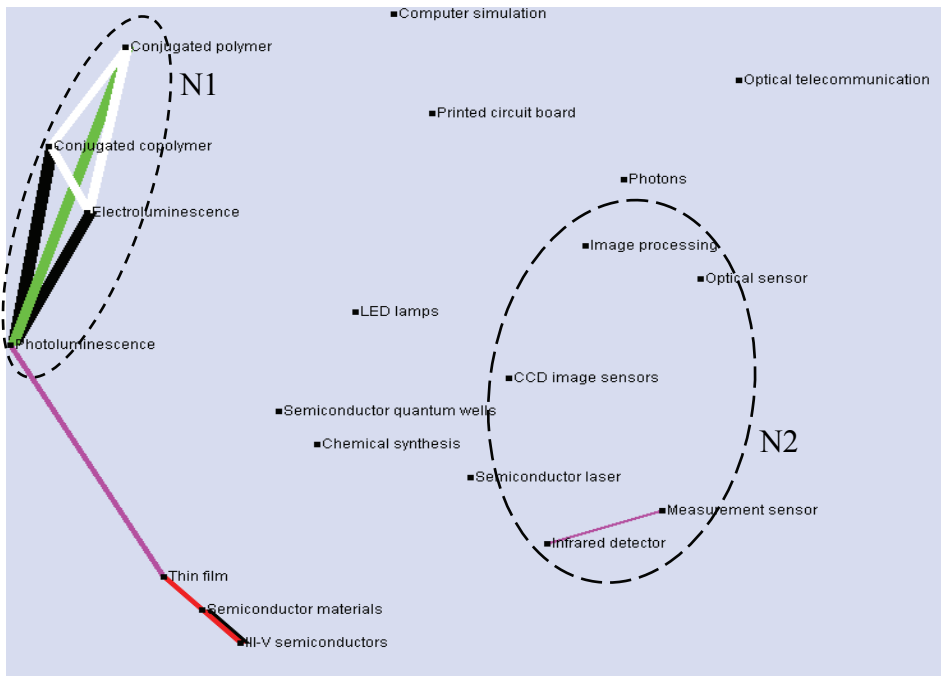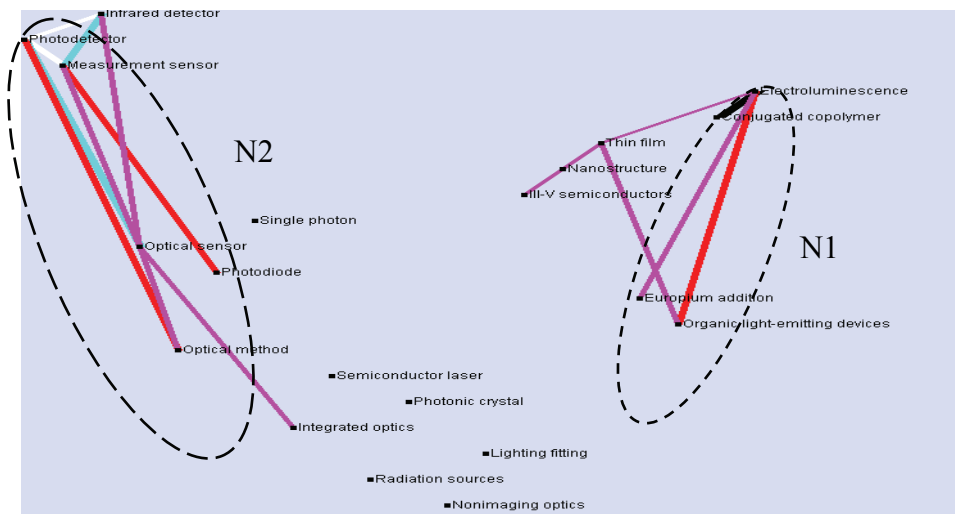
Fig. 11. Networks of clusters for P1

Fig. 12. Networks of clusters for P2

Now we study the complementarity of both methodologies by comparing their results.

5. Convergence between the two methodologies

Once applied to a corpus of 8,169 PASCAL bibliographic records related to optoelectronic devices and divided into two successive periods (2000 to 2004 and 2005 to 2010), both methodologies lead to complementary and converging results.

The stability of inorganic semiconductor materials over the two periods is illustrated by the crossover from the stage "established term" to the stage "cross-section terms" (Pathway V) of keywords like "*III-V semiconductors*" or "*Gallium arsenides*" (diffusion model) and by the diachronic stability of the cluster "*III-V semiconductors*" (cluster analysis).

The leading role of organic materials, and especially organic polymers, over the two periods is highlighted by the presence of keywords like "*Conjugated copolymer*", "*Conjugated polymer*" or "*Phenylenevinylene derivative polymer*" in the "established terms" and "cross-section terms" stages of both periods (diffusion model) and by the strength of the cluster subnetwork N1 right from the first period (cluster analysis).

The growing importance of nanotechnology and nanomaterials in P2 is highlighted by the crossover from the stage "unusual term" to the stage "cross-section terms" (Pathway III) of the keyword "*Nanoparticle*" (diffusion model) and by the emergent role of cluster "Nanostructure" in the second period map (cluster analysis).

The growing importance of imaging devices in P2 is highlighted by the crossover from the stage "unusual term" to the stage "cross-section terms" (Pathway III) of the keyword "*Imager*" (diffusion model) and by the strength of the cluster subnetwork N2 in the second period map (cluster analysis).

The emergence in the second period of applications of optical sensors towards biomedical imaging, highlighted by cluster analysis, is illustrated by the appearance of the keyword "*Biomedical imaging*" amongst the top 10 "unusual terms" of the diffusion model in P2.

The reinforcement of the applications of LED as light sources, highlighted by cluster analysis, is illustrated by the appearance of the keyword "*Daylight*" amongst the top 10 "unusual terms" of the diffusion model in P2. This keyword is related to studies focused on the design of daylight emitting LED lamps.

6. Conclusion

It is a challenge to identify emerging technologies. "Optoelectronic devices" is a broad field and by applying the complementary methods presented here, it can be better characterised and described.

One particularity of our approach is the alternate utilisation of different bibliometric and/or informetric methods and scientific expertise. This expertise is necessary to validate or to complete the results obtained at each step of the work as well as to get the experts personal input on the matter at hand. This could be time consuming, but with our approach, the amount of data submitted to the experts' appreciation is limited, thus making their task notably easier.

The diachronic approach we adopted consists in splitting the corpus in two periods, applying on the one hand the diffusion model to explore the terminological evolution of the subject field, and on the other hand a content cluster analysis for each period to detect the evolution of the topics by examination of the vocabulary related to the respective clusters.

The application of the diffusion model is a novel bibliometric approach giving a more in-depth view of the considered field. Associating terms with technologies allows the

development of a new interesting analysis methodology based on the notion of terminology diffusion. The indicators we used such as term frequency, relative term frequency and age of terms helped work out the different features of a field. Exploiting these results, we were able to understand the field specifications. The bibliometric filter assigning keywords to three stages gives an insight on the emergence of research topics in a technology. At a first glance it is a formal taxonomy used to disassemble what is called a technology in its "atoms" of research and traces the breakthroughs if they happen. Terms in the unusual phase show potentials of future developments in a technology although it is neither a prognosis, nor a prediction. To discuss the merit of emergences and research potentialities it is important to obtain a validation of the results from experts in the concerned field.

Our cluster analysis approach allows to have a global view of the field landscape at two successive time periods. The analysis of the cluster contents and their relative position on the cluster maps supplies indications about their similarity with respect to their respective associated keywords. The observation of cluster maps allows to detect exceptional topics and interesting topic sub-sets. Here also, the experts' input remains at each step absolutely necessary to validate and position the analysis results in the field context by giving them a scientific foundation.

Applying at the same time cluster analysis and diffusion model allows to confirm the results detected by each method and also to lead to a deeper understanding and characterization of the technological field. Furthermore, the diffusion model approach allows new interpretation of clustering results introducing external term categorizations.

7. Acknowledgement

8. References

Armstrong, J. S., Green, K. C. (2007):
http://www.forecastingprinciples.com/selection_tree.html

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 wokshop VSST* (Veille Stratégique Scientifique et Économique), 30-31 mars 2009, Nancy.

Kopcsa A. & Schiebel E. (1998): Science and technology mapping. A new iteration model for representing multidimensional relationships. *Journal of the American Society for Information Science*, 49 (1), 7–17.

Lancaster F. W. & Lee J. L. (1985): Bibliometric techniques applied to issues management: A case study. *Journal of the American Society for Information Science*, 36(8), 389–397.

Lelu A. (1993): Modèles neuronaux pour l'analyse de données documentaires et textuelles. PhD Dissertation, Université de Paris 6.

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

PROMTECH (2007): Identification and Assessment of Promising Emerging Technological Fields in Europe. Final report, July 2007, AIT, Vienna.

Roche I., Besagni D., François C., Hörlesberger M. & Schiebel E. (2010): Identification and characterisation of technological topics in the field of Molecular Biology. *Scientometrics*, 82(3), 663-676.

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. *12th International Conference on Scientometrics and Infometrics*, ISSI 2009, July 14th-17th, 2009, Rio de Janeiro, Brazil.

Schiebel E., Hörlesberger M., Holste, D. Roche I., François C. & Besagni D. (2010): Identification of emerging technologies through pathway-III analysis, *Proceedings of the 19th International Conference on Management of Technology*. Cairo, Egypt, March 8th-12th, 2010

Schiebel E., Hörlesberger M., Roche I., François C. & Besagni D. (2010): An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices. *Scientometrics*, 83(3), 765-781.