Detecting domain dynamics: Association Rule Extraction and diachronic clustering techniques in support of expertise

Ivana Roche, Maha Ghribi, Nathalie Vedovotto, Claire François, Dominique Besagni, Pascal Cuxac, Dirk Holste, Marianne Hörlesberger, Edgar L. Schiebel

Detecting domain dynamics: Association Rule Extraction and diachronic clustering techniques in support of expertise

Ivana Roche[1], Maha Ghribi[1], Nathalie Vedovotto[1], Claire François[1], Dominique Besagni[1], Pascal Cuxac[1], Dirk Holste[2], Marianne Hörlesberger[2], Edgar Schiebel[2]
(1)INIST-CNRS, 2 allée du Parc de Brabois, 54519 Vandœuvre-lès-Nancy Cedex, France
(2)AIT, Austrian Institute of Technology GmbH, Donau-City-Strasse 1, 1220 Vienna, Austria

1 Introduction

Identifying the evolution trends of a scientific domain can be hugely interesting for scientific research policy makers. The evolution of a scientific domain can be studied by associating clustering techniques, generating a representation of the publication scientific landscape based on its extracted terminology, with a diachronic analysis of clustering results obtained at two different times. This work, developed in the context of a European project, aims to propose an alternative way by producing an assisted diachronic analysis of clustering results decreasing the load of the expertise phase.

2 Data

The data sets have been extracted from PASCAL, a multidisciplinary bibliographic database providing broad multidisciplinary coverage and containing nowadays about 20 million bibliographic records resulting from the analysis of the scientific and technical international literature published predominantly in journals and conference proceedings. Moreover each PASCAL record is indexed, either manually by scientific experts or automatically based on a content analysis, by both keywords and thematic categories from a classification scheme. Our study is based on these indexing keywords and is verified by a scientific expert.

The query operated in this work is done by a scientific expert and focuses on a specific field, namely, "Systems and Communications Engineering: electronic, communication, optical and systems engineering" which gathers topics as systems engineering, automation, microelectronics, communication engineering, signal processing, networking or simulation engineering.

Two corpora of bibliographic records dealing with this field have been extracted, for two publication years: 2000 (referred here after to as period *P1*) and 2009 (referred here after to as period *P2*). The number of elements for these two periods is 20,568 for *P1* and 19,827 for *P2*, and the number of indexing keywords is 21,781 for *P1* and 18,475 for *P2*.

3 Methodology

After verifying and validating the consistency and homogeneity of the corpora and its terminology, allowing examining if they well represent the studied field, a clustering process is applied to both periods. This step aims to map each corpus in clusters of similar records with respect to the indexing keywords existing in the bibliographic references forming the corpus. Metaphorically we consider that the obtained cluster maps are a representation of the scientific landscape corresponding to the corpus contents at the two studied time periods.

Our clustering tool 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 ([1], [2]). This step is realized by employing an in-house software tool, Stanalyst [3], devoted to the information analysis.

A diachronic analysis of the clustering results is then operated by assessing the relationships between clusters of the two periods, employing the association rules, classical or fuzzy ([4], [5], [6]), through the so-called "*confidence index*". The goal is firstly to determine which the clusters potentially carrying innovative topics are and to class the set of clusters by rank of innovativeness, given by a so-called Novelty Index. Secondly, we apply a methodology allowing to evaluate the innovativeness degree of new elements by considering their similarity with respect to the clusters with a high innovativeness rank.

We validate this methodology at different steps by means of a huge expertise task consisting, on one hand, on examining the content of each cluster and its relative position in the cluster networks of each period and, on the other hand, on validating both the clusters' innovation ranking and the similarity results calculated for the considered set of new elements.

3.1 The association rules

The association rules are mainly used in frequent patterns mining. They help in finding interesting associations and relationships between item sets in a given data sets. The Market Basket analysis is a typical example for the frequent patterns mining ([4], [5]). The association rules can also help in different data mining tasks such as data classification and clustering.

Let $I = \{I_1, I_2, ..., I_n\}$ be a set of items. An association rule is an implication of the form $A \Rightarrow B$ where $A \subset I$ and $B \subset I$. Two indexes are then calculated for every potential association rule: its "*support*" and its "*confidence*".

The support is defined as the percentage of items that appear in both A and B item sets:

$$support(A \Rightarrow B) = P(A \cup B)$$

This operation has the commutative property: $support(A \Rightarrow B) = support(B \Rightarrow A)$

The confidence is given by the percentage of items that appear in B under the condition that they appear also in A:

$$confidence(A \Rightarrow B) = P(B \mid A)$$

This operation has not the commutative property: $confidence(A \Rightarrow B) \neq confidence(B \Rightarrow A)$

We can then calculate the confidence of $A \Rightarrow B$ by using the support as follow:

$$confidence(A \Rightarrow B) = \frac{support(A \Rightarrow B)}{support(A)}$$

In the context of this work, the items are the keywords (*Kw*) and the item sets A and B are the clusters. We give to a keyword the value 1 if it appears in the item set and 0 if it is absent.

Then, the $support(A \Rightarrow B)$ is the percentage of keywords that appear in A as well as in B and the $confidence(A \Rightarrow B)$ is the percentage of keywords that appear in B under the condition that they appear also in A. The graphical representation of the $support(A \Rightarrow B)$ is presented in the Figure 1.

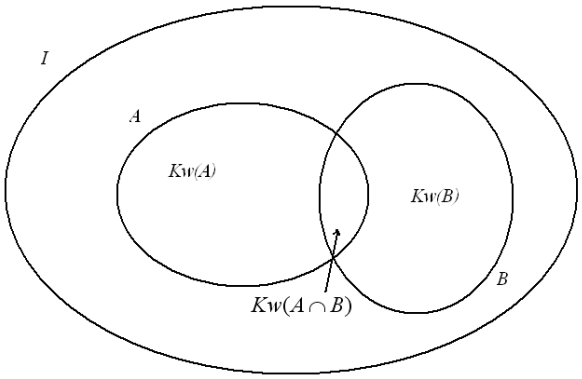

Figure 1: *Illustration of* $A \Rightarrow B$

We calculate:

$$support(A \Rightarrow B) = \frac{Kw(A \cap B)}{card(I)}$$

$$confidence(A \Rightarrow B) = \frac{Kw(A \cap B)}{Kw(A)}$$

The association rule $A \Rightarrow B$ in this context could be interpreted as how much we could consider that the class A is included in B. A value of $confidence(A \Rightarrow B) = 1$ means that all the keywords in A are in B and therefore that A is totally included in B.

In case the appearance of an item in an item set is not evaluated by a binary value, the fuzzy association rules are then used [6]. In the context of our work, the usually considered value is the obtained weight for each keyword in each item set after the clustering step.

The calculation of the $support(A \Rightarrow B)$ is done by using the simple operation of intersection for the fuzzy sets. Thus, for a keyword '*i*' having the value a_i in A and b_i in B, its value in $(A \cap B)$ is equal to $\min(a_i, b_i)$. The Table 1 gives two examples of how to calculate the support and confidence indexes in both cases classical and fuzzy association rules.

***Table 1**: Two examples illustrating how to evaluate the association rule $A \Rightarrow B$ in both cases, classical (a) and fuzzy (b) association rules*

Keywords \ Clusters	A	B	AB
MC1	1	0	0
MC2	0	1	0
MC3	0	0	0
MC4	1	1	1
MC5	1	1	1
MC6	0	1	0
MC7	1	1	1
Total	4	5	3
$support(A \Rightarrow B)$	3		
$confidence(A \Rightarrow B)$	3/4		

a-: Classical Association Rules

Keywords \ Clusters	A	B	AB
MC1	0,6	0,3	0,3
MC2	0,2	0,9	0,2
MC3	0,4	0,9	0,4
MC4	0,9	0,5	0,5
MC5	1	1	1
MC6	0,1	0,9	0,1
MC7	0,8	0,5	0,5
Total	4	5	3
$support(A \Rightarrow B)$	3		
$confidence(A \Rightarrow B)$	3/4		

b-: Fuzzy Association Rules

3.2 A Novelty Index calculation using the association rules

The clustering process applied to the obtained corpus of bibliographic references extracted for 2 publication years (2000 and 2009) produces two sets of clusters. The goal is to sort the clusters of the most recent period from the most to the less innovative on the basis of a diachronic analysis of the clustering results realized by evaluating the relationships between the clusters of the two periods. For that, we develop a Novelty Index (in short *NoI*) as a basis to evaluate the clusters' innovativeness degree by taking into account the evolution of the research developments over the time. This continuity in the time factor will help us to distinguish the emerging topics from the declining ones. We define the *NoI* as a measure of the relationships between the clusters from two periods, named *P1* and *P2*, by using the association rules. We use the fuzzy association rules because our items, namely the keywords of the clusters resulting from the clustering previous step, have non-binary weight values.

Logically, the relationships between two clusters which are considered as close to each other have high confidence indexes. Thus, an innovative cluster of the second period must show small confidence index with regard to each cluster of the first period. Moreover, a class with a topic already introduced in the previous period that keeps developing in the second period could also be considered as innovative but not with the same degree. The clusters that just cover the same topic as a cluster from the previous period is not considered as innovative, even if the topic still interests the researchers. Generally, these clusters are strongly linked to the previous period through one or more clusters.

Considering only the direct relationships between the clusters of the second period with those of the first one could generate a loss of information while reducing its global relationship with the first period. It is for that reason that, in this work, we calculated two different indexes.

The first one measures, for each cluster of the second period, the minimum confidence value among its relationships with each cluster of the first period. It thus evaluates the direct relationship between the two periods. We call it Inter-Period, or *InterP*, because the comparison is realized between the cluster sets of the two periods.

The other index is called Intra-Period, or *IntraP*, because it takes into account the comparison exclusively between clusters from the second period. It allows us to verify on the one hand whether these clusters are strongly linked together and on the other hand if they have potential indirect relationships with the first period, which would not have been detected with *InterP*. The Figure 2 illustrates both, the direct and indirect relationships between the clusters of the second period (*P2*) and those of the first period (*P1*).

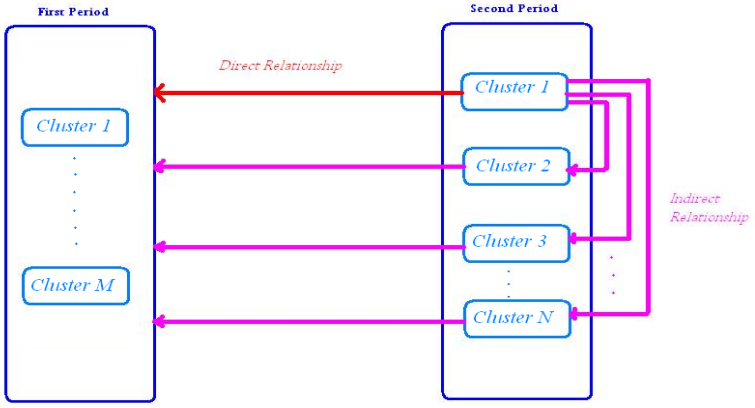

Figure 2: *Illustration of the two types of cluster relationships between the two periods*

3.2.1 *The Inter-Period index*

This index considers exclusively the direct relationships between the clusters of the second period and those of the first period. For each cluster *i* from the second period we define *InterP* as follows:

$$InterP_i = \max_{j \in P1} \{Cf(i \Rightarrow j)\}$$

where:
P1 represents the set of clusters of the first period;
$Cf(i \Rightarrow j)$ represents the value of the confidence index of the association rule $(i \Rightarrow j)$.

This index calculates the maximum value of the linkage of this cluster with all clusters of the previous period. The lower the value of *InterP,* the stronger the innovativeness degree of the cluster.

3.2.2 *The Intra-Period index*

This index must allow to answer two questions:
- how much strongly is each cluster *i* of the second period linked with the other clusters of the same period?
- is it highly linked to the clusters of the first period? Thus we should be able to identify whether there are potential indirect relationships between the considered *i* cluster and the *P1*'s clusters, that were not identified by the only calculation of InterP.

As a first idea, for every cluster *i* from the second period, we look for the clusters from the same period, which are highly linked with *i*.

Let C_i be the set of clusters from *P2* that have a value of confidence index with the cluster *i* higher than a threshold δ fixed manually:

$$C_i = \{ j \in P2, Cf(i \Rightarrow j) \geq \delta \}$$

The $IntraP_i(\delta)$ is then defined as the mean of the *IntraP* of the clusters of C_i and calculated as follows:

$$IntraP_i(\delta) = \frac{1}{|C_i|} \sum_{j \in C_i} InterP_j$$

The value of *NoI* could be then calculated as the mean of the *IntraP* and the *InterP* and, moreover, these mean values could allow classing the clusters of the second period by their rank of innovativeness.

Nevertheless we noticed that the choice of the value of the threshold δ is a very big disadvantage of this method. Indeed we observed that, in some cases, even a very little change of its value could change significantly the result namely the order of the clusters in the innovativeness ranking. In fact, we examined the behaviour of this threshold in real cases and we found a too important instability in the order of clusters we obtained while changing its value.

So the idea to avoid this threshold is to consider all the clusters of *P2* to calculate *IntraP*. The problem lies in the fact that the importance of every cluster varies with the value of its confidence index with the cluster *i*. That means that the clusters which are highly linked to *i* are very important for us whereas those which are weakly linked to *i* are not. To resolve this question we introduce a weighting function which takes into account the importance of the participation of clusters in *IntraP*.

Thus, we are going to divide the interval $[0;1]$ into 10 sub-intervals defined as follows:

$$In_k = [0.1k; 0.1(k+1)] \text{, with } k = 0,...,9$$

Then, for each cluster *i*, and for every sub-interval In_k, we calculate:

$$IntraP_i(In_k) = \frac{1}{\left|C_i^k\right|} \sum_{j \in C_i^k} InterP_j$$

where C_i^k is the set of clusters from *P2* that have a value of confidence index with the cluster *i* in the sub-interval In_k:

$$C_i^k = \left\{ j \in P2, Cf(i,j) \in In_k \right\}$$

The weighting function w_g is developed so that, being given two sub-intervals In_k and In_l ($k,l \in \{1,...,9\}$), if *k*<*l* then $w_g(In_k) < w_g(In_l)$.

We define then the following increasing weighting function:

$$w_g(In_k) = \frac{1}{10-k};\ k = 0,...,9$$

With this condition, we make all the confidence index values that belong to the upper sub-intervals more important than the others in calculating $IntraP_i$.

The index $IntraP_i$ is then calculated as the weighted mean of the $IntraP_i(In_k)$ as follows:

$$IntraP_i = \sum_{k=0,...,9} w_g(In_k) IntraP_i(In_k)$$

3.2.3 *The Novelty Index*

The global value of the Novelty Index is defined as the harmonic mean of the *IntraP* and the *InterP* indexes. Thus, the lower the cluster's *NoI* value, the higher its innovativeness degree or, in other words, the more it carries positive dynamic changes. Indeed, a *P2*'s cluster with an index *NoI* near to the zero value means that both, its *IntraP* and its *InterP*, are low. This cluster is weakly linked, directly and indirectly, to the clusters from the first period and the keywords representing it deal with topics potentially new.

3.3 Novelty of a new element

In the previous section we described the process bringing us to obtain a Novelty Index value for each *P2*'s cluster. We are now interested on determining the innovativeness degree of any new element with regard to the *P2*'s cluster map that, let us remind, represents the most recent scientific landscape of the studied domain.

In a first step, we apply a text mining approach to extract from any considered new element the terminological information allowing to get a characterization as discriminant as possible in order to represent its content as faithfully as possible. Each new element is then represented by a binary vector showing the presence of its indexing keywords by the value 1 and 0 otherwise. Finally, our methodology associates, to any new element, the *P*2's clusters to which it is the most similar and determines, from this information, its innovativeness degree.

Evaluating the *NoI* index for the *P*2's clusters and sorting them from the most to the less innovative is a good basis to measure the innovativeness of a new element. We can indeed consider that the closest the new element is to clusters of positive dynamic changes, the more innovative it is. But the vectors representing on the one hand the content of a cluster and on the other hand a new element are formed by numerical values of different types.

For each cluster, the employed classification method calculates to each one of its keywords a real numerical value that assesses how much the cluster could be described by this keyword: we call it the keyword "*weight*" in the considered cluster. So each cluster is represented by a non-binary vector, while each new element is represented by a binary one. Therefore, neither the Euclidian distance nor the cosine similarity is very useful to calculate the proximity between the new elements and the clusters. The idea is then to assign to the proposal the cluster whose keywords represent it at best.

We could, for instance, calculate, for each cluster, the mean of the weights of the keywords that appear in the indexing of the new element as well as in the cluster. The new element would be then assigned to the clusters getting the highest values. But this approach does not take into account the distribution of the keywords in the cluster. Thus, instead of using directly the keyword's weights we calculate with which probability each keyword could be considered as important relatively to the distribution of the keywords indexing the new element in the cluster. We evaluate the cumulative distribution function (*CDF*) corresponding to the weight values of the new element's keywords in the considered cluster.

Let us call W_i the variable that takes as value the weight of a keyword in a cluster *i*. For any value w, we calculate the corresponding cumulative distribution function value as follows:

$$F_{W_i}^{i}(w) = \int_{-\infty}^{w} f_{W_i}^{i}(u)du = P[W_i \leq w]$$

where $f_{W_i}^{i}$ is the density function of W_i.

Theoretically, $F_{W_i}^{i}(w)$ is the probability that the observed value of W_i will be at most equal to w. It can be also regarded as the proportion of the keywords whose weight is lower to w. If $F_{W_i}^{i}(w)$ is near to *1*, this means that the keyword is highly significant in this

cluster and represents it well. Conversely, if $F_{W_i}^i(w)$ is far from *1*, this means that the keyword is not very important in this cluster because there are other keywords that have weights higher than w. In fact, if almost all the keywords have a weight less than w this means that it is one of the most important weights in this cluster.

The similarity value between a new element and a cluster is then calculated as the mean of the values of the *CDF* of the keywords that appear in the new element as well as in the cluster:

$$Similarity(n,i) = \frac{1}{|W_n|} \sum_{w \in W_n} F_{W_i}^i(w)$$

where:
n represents the new element;
i represents the cluster and
W_n is the set of weight values of the new element's keywords in the cluster.

The new element is then assigned to the clusters with which it gets the highest similarity values. The interpretation of these results is quite easy: the lower the *NoI* value of these clusters, the stronger the innovativeness degree of the new element.

4 Results and discussion

The clustering process applied to the keywords of each corpus leads to cluster maps representing the landscape of the field "Systems and Communications Engineering" for each time period. We present in the Figure 3 the obtained map of the 50 clusters of *P2*.

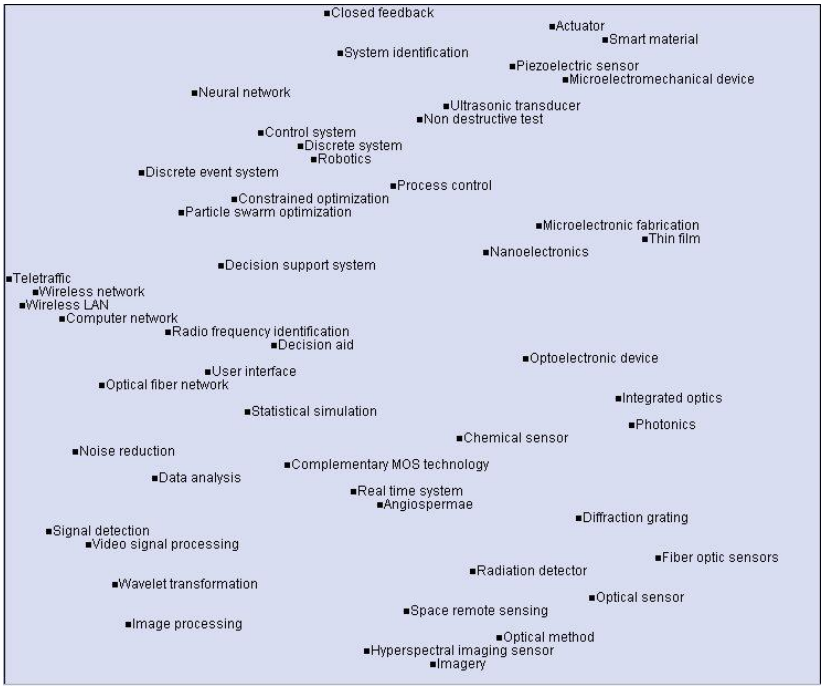

Figure 3: *Obtained cluster map for P2*

The Novelty index of each cluster of *P2* is then calculated, leading to a classification of those clusters according to their innovativeness with regards to *P1*. This innovativeness is qualified as "high", "intermediate" or "low".

***Table 2**: Distribution of P2 clusters innovativeness according to their Novelty index*

High	Intermediate	Low
Angiospermae	Optical method	Decision support system
Space remote sensing	Thin film	Optoelectronic device
Statistical simulation	Nanoelectronics	Imagery
Decision aid	Non destructive test	Image processing
Radio frequency identification	Chemical sensor	Computer network
Complementary MOS technology	Smart material	Closed feedback
Data analysis	Microelectromechanical device	System identification
Discrete event system	Wavelet transformation	Photonics
Discrete system	Neural network	Fiber optic sensors
Process control	Particle swarm optimization	Wireless network
Ultrasonic transducer	User interface	Optical fiber network
Control system	Optical sensor	Integrated optics
Hyperspectral imaging sensor	Video signal processing	Signal detection
Microelectronic fabrication	Piezoelectric sensor	Teletraffic
Real time system	Constrained optimization	Wireless LAN
Radiation detector	Actuator	Diffraction grating
	Robotics	
	Noise reduction	

These results are consistent with those formerly obtained by the expert by the study of the comparison matrix ([7], [8]) built with the vocabularies of the two periods by giving the fraction of keywords belonging to the second period clusters that already appear in the P1's clusters. Indeed, several *P2*'s scientific themes have been identified as innovative by the two methodologies, as remote sensing, decision aid applications to medicine or biotechnologies. Furthermore, the approach based on association rules reinforces the highlighting of innovative themes, as radio-frequency identification or hyperspectral imaging. We thus observe a real convergence between the results obtained with the developed approach, which brings a first validation step to our method.

The set of new elements that we employ in this work is a sample of 29 project proposals submitted for funding in 2009 and dealing with the considered scientific field. They are represented by their respective title and abstract and are tagged with a binary value, namely 1 if the proposal has been accepted (4 proposals in our sample) and 0 otherwise. This tagging was done by a scientific expert panel that, after examining all the submitted proposals, has made a choice based on the four principal criteria of selection enounced by the ERC (European Research Council). They allow

characterizing each project proposal according to its innovativeness, applicability, risk and interdisciplinarity [9].

It is important to notice that, in this work, we develop a model taking into account only a part of the selection process done by the expert panel, namely the evaluation of the innovativeness degree of each proposal. Consequently, the results we obtain can be different from those coming from the expert panel, which considered the whole set of the ERC's selection criteria.

A text mining step, based on the terminological informations present in the titles and abstracts of our sample, allows to represent each project proposal by a binary vector. The calculation of the similarity (cf. section 3.3) allows us to generate, for each project proposal, a set of 50 values calculated by considering its similarity with each one of the *P2*'s clusters. Among these values we consider on the one hand the 4 first highest, corresponding to the 4 clusters with which the project has the highest similarity and on the other hand the calculated *NoI* value of each of these 4 clusters.

Thus, by combining these two data, we produce a classification of the 29 project proposals by decreasing rank of the value of their innovativeness degree. This value is calculated according to the classification of the innovativeness degree of the *P2*'s clusters presented in Table 2.

The comparison of the obtained results with the choice performed by the expert panel shows that 2 out of the 4 accepted project proposals are present in the first third of our classification by decreasing innovativeness degree. Table 3 presents these Top10 and indicates for each one the choice of the expert panel.

***Table 3**: List of the Top10 project proposals by decreasing rank of innovativeness degree*

Project proposal ID	Innovativeness degree rank	Expert panel choice (0/1)
PROP_19	1	0
PROP_23	2	0
PROP_14	3	0
PROP_02	4	0
PROP_08	5	**1**
PROP_07	6	0
PROP_22	7	0
PROP_06	8	0
PROP_12	9	0
PROP_01	10	**1**

If 2 out of the 4 project proposals chosen by the expert panel are in the Top10 list, the two other ones are classified respectively at the 20th and 29th position of our classification of innovativeness degree. This difference between the experts' assessment and our results could have many causes.

First of all, our methodology takes into account only 1 out of the 4 criteria of selection followed by the expert panel, and these 2 proposals could thus have been chosen due

to an important value of one or more of the 3 other criteria. Secondly, the terminological referential employed in the text mining step could miss the most recent terminological contributions of the studied scientific field. Thirdly, we calculate the innovativeness degree of each proposal only considering the first 4 higher similarity values instead of the 50 available.

As far as the first point is concerned, we are currently working in the framework of an European project to the development of a modelling of the three other ERC's selection criteria, with the goal to produce a composite indicator that a priori should better reflect the effects of the complete set of criteria applied by the expert panel.

The second point is a little bit delicate. Indeed, we would have been able to update our representation of the landscape of the considered field by employing, in our diachronic analysis, a *P2* more recent than 2009 and containing the most recent terminological information related to the field. But in this case we risk to introduce a bias because new knowledge appearing in 2010 were not still known in 2009. Furthermore, the goal is, in fine, to apply our methodology concurrently to the expert panel, if not before, and surely not one year later.

Finally, we can consider in the calculation of the innovativeness of each proposal a number of calculated values of similarity larger than 4, even the whole set (in this work equal to 50). The sensibility of the indicator to the variations of the number of similarity values effectively considered in its calculation will be studied. Furthermore, it could turn out also useful to help us to analyze the case of the project propositions getting only similarities with the clusters having low values of *NoI*.

Acknowledgements
Methodology is developing in the context of the DBF (**D**evelopment and Verification of a **B**ibliometric Model for the Identification of **F**rontier Research) project within the Coordination and Support Actions (CSAs) of the IDEAS specific programme of the EU's 7th Framework Programme for Research and Technological Development (project reference n° 240765). The authors wish to acknowledge this support.

5 References

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

[2] A. Lelu and C. François (1992), "Hypertext paradigm in the field of information retrieval: A neural approach", 4th ACM Conference on Hypertext, Milano, November 30th–December 4th

[3] X. Polanco, C. François, J. Royauté, D. Besagni, I. Roche (2001), "Stanalyst®: An integrated environment for clustering and mapping analysis on science and technology" In: Proceedings of the 8th ISSI, Sydney, July 16th -20th

[4] J. Han and M. Kamber (2001), "Data Mining : Concepts and Techniques", San Francisco : Morgan Kaufmann Publishers

[5] D. Hand, H. Mannila and P. Smyth (2001), "Principals of Data Mining", Cambridge, Massachusetts, USA: The MIT Press, 2001

[6] P. Cuxac, M. Cadot and C. François (2005), "Analyse comparative de classification: Apport des règles d'association Floue" In EGC 2005, pp. 519–530

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

[8] I. Roche, N. Vedovotto, D. Besagni, C. François, R. Mounet, E. Schiebel, M. Hörlesberger (2011), "Identification of emergent research issues: the case of optoelectronic devices", Optoelectronic Devices and Properties, Oleg Sergiyenko (Ed.), ISBN: 978-953-307-204-3, InTech Available from: http://www.intechopen.com/articles/show/title/identification-of-emergent-research-issues-the-case-of-optoelectronic-devices

[9] D. Holste, I. Roche, M. Hörlesberger, D. Besagni, T. Scherngell, C. François, P. Cuxac, E. Schiebel (2011), "A concept for inferring "Frontier Research" in research project proposals", in: Proceedings of the 13th ISSI, Durban, July 4th – 7th, pp. 315-326