

AMELIORER LA VEILLE DES CHERCHEURS DANS UN CONTEXTE ACADEMIQUE : LES APPORTS CONJOINTS D'UN OUTIL DE VEILLE, DE RESSOURCES HUMAINES ET DOCUMENTAIRES D'UNE BIBLIOTHEQUE DE RECHERCHE.

Stéphanie SAVINA (*), Jean-Michel CAREIL (**), Luc GRIVEL (***)

Stephanie.savina@enscp.fr, jmcareil@intellixir.com, luc.grivel@univ-paris1.fr

(*) Univ. Paris-Est Marne-la-Vallée, Science et Ingénierie de l'Information et de l'Intelligence Stratégique, 2 allée du Promontoire, 93160 Noisy-Le-Grand, France

(**) société Intellixir, 2 boulevard de la Plaine, 04100 Manosque, France

(***) Univ. Paris I, 17 rue de la Sorbonne 75005 Paris, France, chercheur associé à l'équipe S3IS, Univ. Paris-Est Marne-la-Vallée, 2 allée du Promontoire, 93160 Noisy-Le Grand, France

Mots clefs :

Outil de veille, Appropriation outil de veille, documentaliste, analyse de données

Keywords :

Competitive intelligence software, Information scientist, competitive intelligence, data analysis

Palabras clave :

Softwares de inteligencia competitiva, analisis dato

Résumé

Sur la base d'une expérience menée à l'ENSCP (Ecole Nationale Supérieure de Chimie de Paris), cet article identifie **2 facteurs qui conduisent à une** amélioration de la veille d'un chercheur dans un laboratoire de recherche disposant d'accès aux ressources électroniques usuelles et notamment aux bases de données en ligne:

- l'emploi d'un outil proposant des capacités de traitement des corpus de données bibliographiques selon des axes personnalisés d'analyse de l'information, à savoir suivant les centres d'intérêt du chercheur ;
- le rôle d'un professionnel de l'information (bibliothécaire, documentaliste ...) garant de la qualité de l'information et des traitements appliqués.

Il montre également l'intérêt dans un projet de recherche d'un **modèle d'organisation dual** avec, d'une part un centre de documentation négociant les accès aux sources d'information et disposant de personnels maîtrisant les outils de veille, présent dans les laboratoires, et de l'autre des réseaux d'experts (chercheurs) compétents dans l'analyse de l'information.

Abstract

Based on experiments conducted at the ENSCP (technical university and research laboratories), this article identifies 2 key factors contributing to improve the competitive intelligence of a researcher having access to online journals and databases:

- The tool capacity to deal with large corpus of information. More specifically its ability to analyse the bibliographic references corpus according to different aspects of the researcher's interest that has been previously identified.
- Active involvement of the information scientist in increasing the quality of the information research process.

Furthermore close collaboration between researches and library team and availability of high quality library infrastructure is essential in succeeding the research process efficiently. Whereas the researchers perform information analysis, the library department provides online journals and databases subscription, and makes available information scientists skilled in competitive intelligence tools. These experts work along with the research team.

1 Introduction

Aujourd'hui dans un contexte de compétition accrue entre laboratoires, face à une pression des organismes de recherche pour la valorisation, se pose à tout directeur de recherche la question de l'**optimisation de l'usage des ressources documentaires**, du point de vue **de l'argent investi, du temps consacré à la recherche, du traitement de l'information collectée**, ainsi que de **la pertinence de l'information trouvée**.

L'objectif de cet article est d'identifier les **facteurs qui vont contribuer** à améliorer la veille d'un chercheur dans un laboratoire de recherche disposant d'accès aux ressources électroniques usuelles : bases de données, bouquets de titres en ligne.

Les bases de données structurées sont les outils de prédilection du chercheur : identification de nouvelles revues, sélection des articles les plus pertinents, identification de nouveaux axes de recherche, de nouvelles thématiques, de nouveaux partenariats, tels sont les objectifs poursuivis par les chercheurs dans le cadre de leur veille scientifique.

Parallèlement, sont apparus de nouveaux outils de veille, capables de traiter de grands volumes d'informations. Formés à l'utilisation de ce type d'outils, les professionnels de l'information d'aujourd'hui n'ont-ils pas un rôle à jouer en direction du chercheur pour accompagner l'intégration de ce nouveau type d'outil dans les pratiques de recherche ?

L'étude menée ici est basée sur une expérimentation d'utilisation d'un outil de veille dans un contexte de recherche académique. Elle met en synergie les compétences, d'un chercheur (J.F. Guillemoles, CNRS, UMR 7575, Ecole Nationale Supérieure de Chimie de Paris), d'un créateur de logiciel de veille (J.M. Careil, logiciel « Intellixir ») d'une bibliothécaire (S. Savina, ENSCP) et d'un enseignant (L. Grivel, enseignant Univ. Paris I, et Paris-Est Marne-la-Vallée). Cette étude fait suite aux travaux initiés dans le rapport de stage du Master IIDC, soutenu en 2006, à l'université de Marne-la-Vallée, de S. Savina sur la cartographie des réseaux scientifiques.

L'article est structuré de la façon suivante :

- L'analyse des résultats d'une première expérimentation, est décrite dans la section 1. La section 2 sera dédiée d'une part au traitement de corpus de données bibliographiques selon des axes personnalisés d'analyse de l'information, et d'autre part à l'apport possible d'un professionnel de l'information pour favoriser la qualité des résultats obtenus.

2 PHASE 1

2.1 Méthodologie

2.1.1 Exemple de sujet de recherche : « thin layer + photovoltaic »

L'objectif est de suivre l'évolution, de 1992 à 2006, d'une thématique stratégique pour un laboratoire : les couches minces dans le domaine photovoltaïque. Un corpus de 622 notices, constitué de références bibliographiques et références de brevets, est extrait de la base bibliographique «Chemical Abstracts » sous sa forme électronique : « ScienceFinder ». Cette base indexe un grand nombre de journaux, brevets du domaine de la chimie.

L'objectif est de mettre en relief l'information utile au chercheur à partir du traitement de larges corpus de références bibliographiques dont la lecture des notices une à une, est pratiquement impossible car trop chronophage [1]. Nous souhaitons déterminer les concepts principaux relatifs à la thématique des couches minces dans le secteur photovoltaïque.

Le type de données à traiter et l'étude de benchmarking des produits de veille menée par l'INIST [2], nous ont conduits à choisir l'outil « Intellixir » [3] pour effectuer ce traitement.

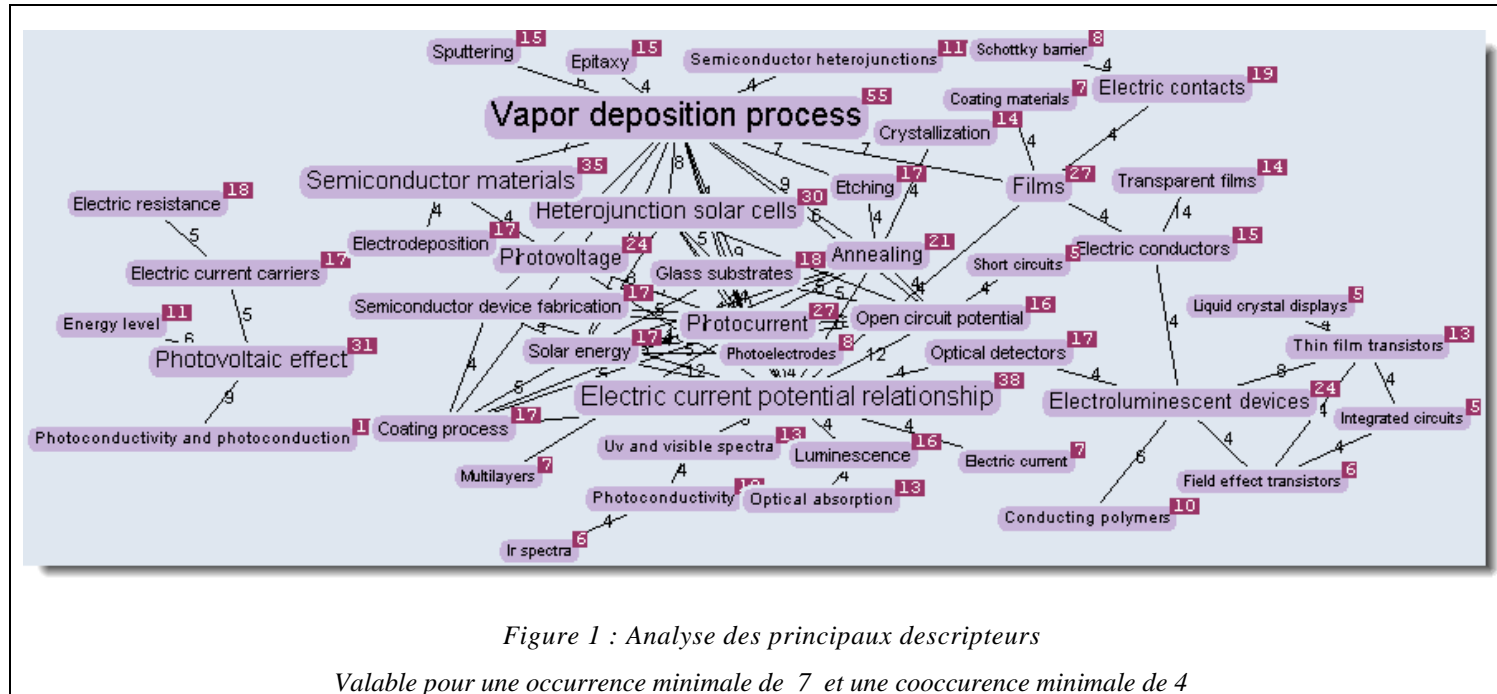
2.1.2 «Intellixir » : mesure

Cette société créée en 2002, commercialise ce Système basé sur des développements issus du Commissariat à l'Energie Atomique. « Intellixir » applique un traitement statistique basé sur la méthode des mots associés [4], qui identifie :

- les principales thématiques de recherche
- les principaux acteurs (académiques et privés) et leurs collaborations les plus fréquentes
- la répartition des publications dans le temps, et entre les pays

Il permet également d'identifier des concepts émergents à partir de la cartographie des réseaux d'auteurs, de laboratoires, et des thèmes de recherche.

A partir des notices de la base « Chemical Abstracts », « Intellixir » extrait une première représentation cartographique (figure 1) à partir du champ « Indexing Terms » de la base. Les « Indexing Terms » sont classés en fonction de leurs occurrences, pour une lecture efficace des graphes, le système propose de limiter l'analyse aux 200 premiers.

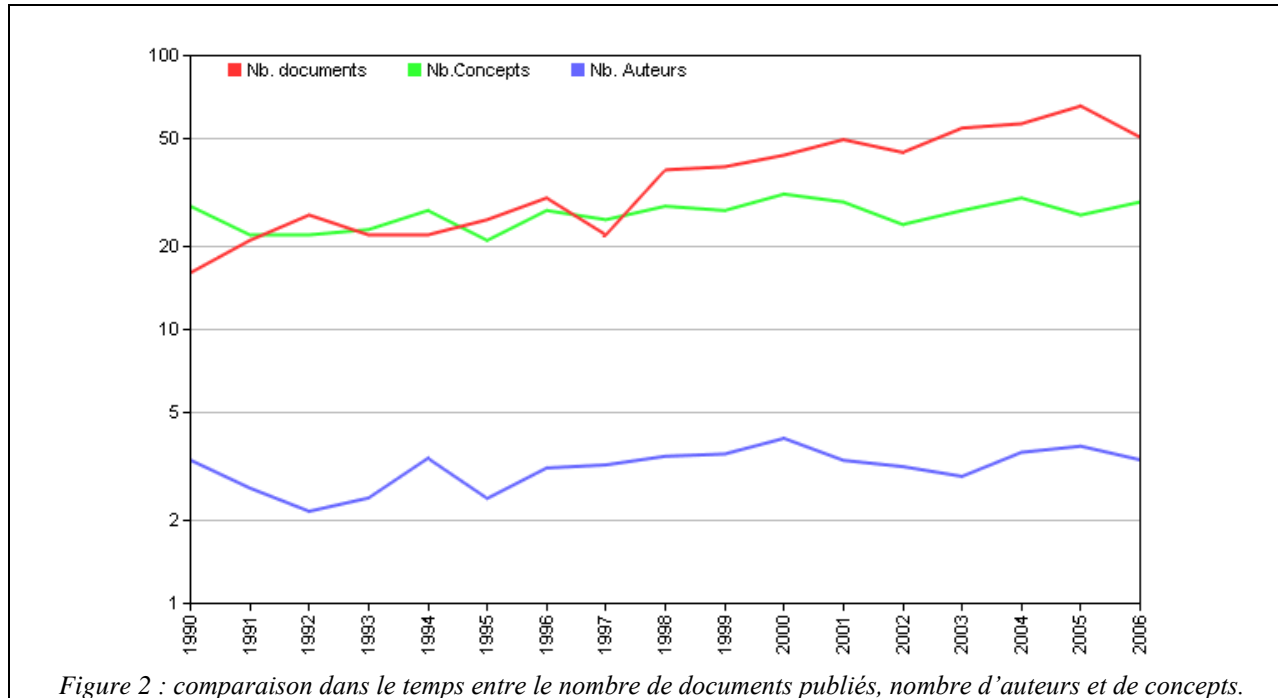


Deux descripteurs i, j sont co-occurent, lorsqu'ils apparaissent simultanément dans un même document. La lecture du graphe se fait de la façon suivante :

- En rouge, le nombre de fois où le terme apparaît dans le corpus
- Le chiffre mentionné sur le lien correspond au nombre de fois où les termes qui sont reliés co-occurent dans le même document.

Il est possible de faire varier simultanément le palier minimum de ces deux critères, ce qui permet de focaliser sur des parties du réseau et/ou obtenir une vue générale des grandes thématiques qui se dégagent.

Cette représentation est susceptible d'intéresser un chercheur lorsqu'il entame une recherche bibliographique sur un nouveau sujet, ou doit présenter une vue d'ensemble sur l'évolution d'une discipline (figure 2) ou bien proposer un outil d'aide à la décision [5], et de gouvernance. Par ailleurs, il est possible de suivre l'évolution des termes par leur fréquence d'apparition sur des périodes de un an par exemple, et par conséquent de détecter les thématiques émergentes de la discipline.



Cependant cette représentation se réfère à une indexation réalisée au travers d'un thésaurus de termes normalisés [6]. Or, il est souvent reproché au documentaliste de poser des mots sur un concept exprimé dans un document et par conséquent ne pas refléter la pensée de l'auteur. Aussi, dans un objectif de recherche pertinente d'informations utiles au chercheur, « Intellixir » propose une autre approche de traitement du corpus à partir de l'analyse des mots du titre et du résumé (figure 3, figure 4) en fonction de leurs occurrences.

Role of defects at nanoscale ZnO and Cu(In,Ga)Se₂ semiconductor interfaces

EXTRACTION



Concepts : NANOSCALE ZNO ; CU ; GA ; SE2 SEMICONDUCTOR INTERFACE ;

Figure 3 : extraction des concepts à partir du titre

Résumé : Current understanding and control of semiconductor contacts increasingly require measurements sensitive to defects and chem. changes at nanoscale interfaces. We offer examples illustrating dramatic macroscopic effects occurring in semiconductor systems as a result of nanoscale interface phenomena. In some cases, there is interplay of several competing defect-driven mechanisms. Elucidating them and finding the leading ones require careful exptl. approach. For single-cryst. ZnO, we study the role of near-surface defects on the formation of Au Schottky contacts. Among the factors degrading the rectifying characteristics of such contacts one should consider the following. High concns. of shallow donors in the surface and subsurface regions lead to barrier thinning, resulting in increased tunneling. Alternatively, the presence of deep defects near contact interface promotes tunneling by defect-assisted hopping. Nanoscale electronic and chem. studies show that independent redn. of both shallow donors and deep defects significantly improves the rectifying performance of the Au/ZnO contacts. We find that processing of ZnO with remote O and H plasmas allows for controllable tailoring of chem. and phys. properties of the surface. By the same token, nanoscale compositional and electrostatic variations between grain boundaries and grain interiors in thin polycryst. films of Cu(In,Ga)Se₂, absorber layers in record-setting solar cells, show how nanoscale arrangement of near-surface stoichiometric defects may improve the overall photovoltaic efficiency. Confirming the theory, we find a 50% redn. in Cu compn. from grain interior to boundary and a p-type potential barrier that acts to reduce majority-carrier hole recombination. These examples emphasize the practical significance of nanoscale chem. and electronic features at electronic material interfaces. [on SciFinder (R)]

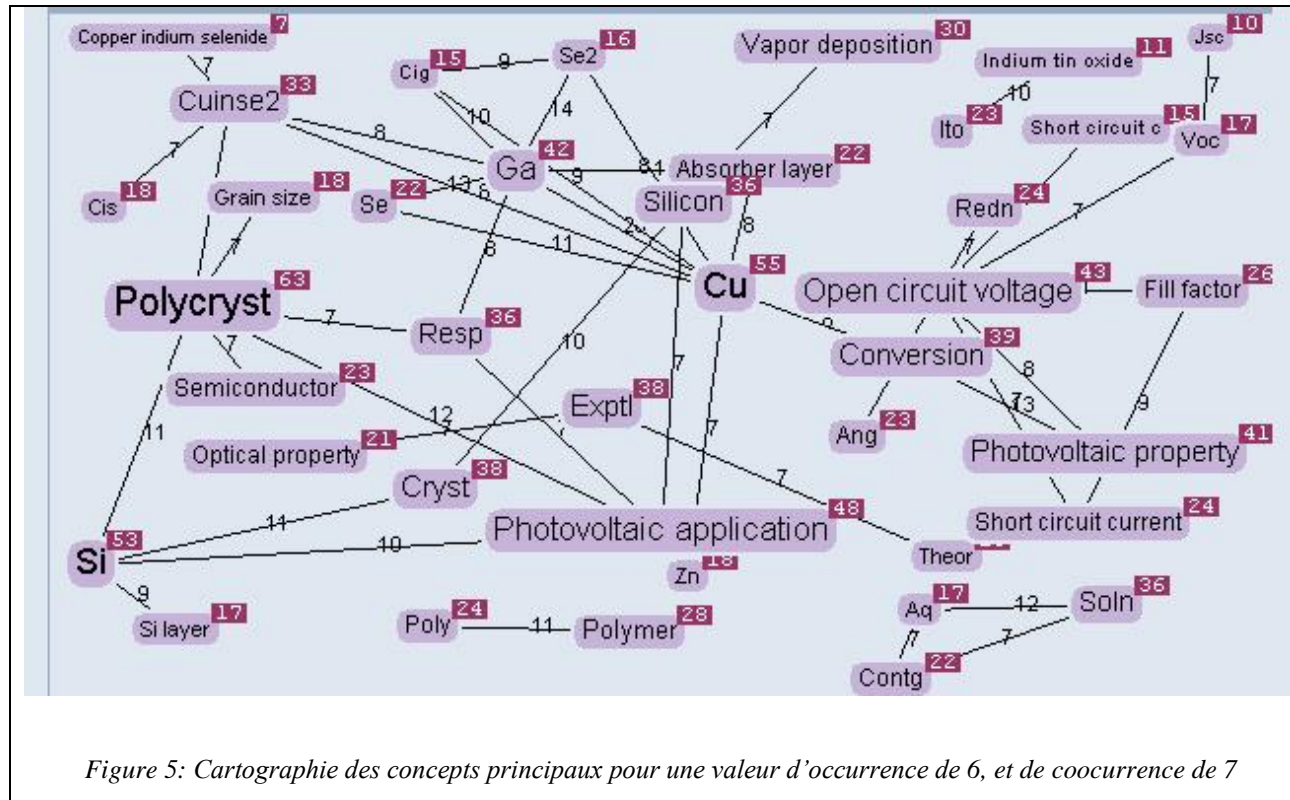
EXTRACTION



Concepts : SEMICONDUCTOR CONTACT INCREASINGLY ; MEASUREMENT SENSITIVE ; CHEM ; NANOSCALE INTERFACE ; DRAMATIC MACROSCOPIC EFFECT OCCURR ; SEMICONDUCTOR SYSTEM ; NANOSCALE INTERFACE PHENOMENA ; INTERPLAY ; COMPETE DEFECT DRIVEN MECHANISM ; CAREFUL EXPTL ; SINGLE CRYST ; ZNO ; NEAR SURFACE DEFECT ; AU SCHOTTKY CONTACT ; FACTOR DEGRAD ; RECTIFY CHARACTERISTIC ; HIGH CONCN ; SHALLOW DONOR ; SUBSURFACE REGION LEAD ; BARRIER THINN ; TUNNEL ; DEEP DEFECT NEAR CONTACT INTERFACE PROMOTE TUNNEL ; DEFECT ASSIST HOPP ; NANOSCALE ELECTRONIC ; CHEM ; INDEPENDENT REDN ; SHALLOW DONOR ; DEEP DEFECT ; RECTIFY ; AU ZNO CONTACT ; ZNO ; REMOTE O ; H PLASMAS ; CONTROLLABLE TAILORE ; CHEM ; PHYS ; TOKEN ; NANOSCALE COMPOSITIONAL ; ELECTROSTATIC VARIATION ; GRAIN BOUNDARY ; GRAIN INTERIOR ; THIN POLYCRYST ; CU ; GA ; SE2 ; ABSORBER LAYER ; RECORD SET SOLAR CELL ; NANOSCALE ARRANGEMENT ; NEAR SURFACE STOICHIOMETRIC DEFECT ; PHOTOVOLTAIC ; REDN ; CU COMPN ; GRAIN INTERIOR ; P TYPE POTENTIAL BARRIER ; MAJORITY CARRIER HOLE RECOMBINATION ; EMPHASIZE ; PRACTICAL SIGNIFICANCE ; NANOSCALE CHEM ; ELECTRONIC FEATURE ; ELECTRONIC MATERIAL INTERFACE ; SCIFINDER ;

Figure 4 : extraction des concepts à partir du résumé

Ces mots extraits sont affectés au champ «concept » et une autre représentation cartographique est construite (figure 5).



Au contraire de la figure 1 qui analysait les descripteurs, indexés par « Chemical Abstracts », apparaissent cette fois-ci, les noms de matériaux (Cu, GA, CuInS2) et des techniques. Cette représentation se rapproche de la représentation mentale du chercheur en quête d'informations sur l'action d'un produit dans un contexte d'expérience. Cependant, la cartographie des « concepts » soumise au chercheur est difficilement exploitable car un travail de normalisation des concepts extraits est nécessaire. Comme préalable, il est obligatoire de lier les expressions synonymes « Si » et « silicon », traiter les abréviations de la base « Chemical Abstracts », « poly » et « polymers ». L'import du dictionnaire des abréviations des « Chemical Abstracts » permet d'automatiser cette étape.

Notre objectif qui est d'offrir au chercheur la possibilité d'avoir une lecture d'un corpus de grande densité est certes atteint, mais le chercheur n'a été que partiellement satisfait par les résultats obtenus, pour des causes qui ont peu à voir avec l'outil utilisé, mais beaucoup avec la démarche qui a été suivie.

2.2 Interprétation des résultats : Analyse de la performance

2.2.1 Nécessité d'une couverture exhaustive des publications du laboratoire

La base « Chemical Abstracts » est souvent considérée comme source d'information incontournable, nécessaire et suffisante. Mais il est apparu qu'une partie des publications du laboratoire se situe à la frontière entre les sciences des matériaux et la chimie et n'apparaît ainsi pas dans le corpus retenu. Le chercheur doute de la fiabilité de la recherche et donc de la pertinence de l'analyse des résultats produits par l'outil de veille.

Dans le cadre de ce test, il aurait plus été judicieux de couvrir l'exhaustivité des publications du laboratoire et d'étendre la couverture à des bases spécialisées dans les matériaux.

2.2.2 Usage ponctuel

Le test d'un outil de quelque nature qu'il soit est forcément ponctuel, du temps et un investissement sont nécessaires pour se familiariser avec l'outil, et le paramétrer. Ces opérations sont malheureusement ressenties comme chronophage au regard de l'absence de perception immédiate des capacités du produit. Ce qui entraîne manque de motivation et manque de capitalisation.

La seconde partie est relative à la présentation d'une méthodologie qui doit optimiser la qualité de la recherche d'information et sa pertinence en personnalisant le corpus tout en réduisant le temps utilisé au paramétrage de l'outil.

3 Phase II

3.1 Création de champs personnalisés

J M Careil propose une approche spécifique pour répondre aux attentes des chercheurs. Il propose de personnaliser le corpus non plus en fonction des descripteurs fournis par Chemical Abstract ou tout autre producteur de bases de données, mais dorénavant paramétrer l'outil à partir des connaissances du chercheur. Ce dernier **formalise les concepts à extraire du corpus**, en collaboration avec un spécialiste de l'outil (veilleur, documentaliste, gestionnaire de l'information).

La démarche est la suivante :

- 1) Extraction automatique des concepts pour chaque notice sur les champs titre et résumé (figure 3 et 4)
- 2) Sélection aléatoirement d'une dizaine de notices

- 3) Analyse de chaque notice. Dialogue entre le chercheur et le spécialiste pour identifier les thématiques, les mots clés, les possibles synonymes utilisés pour décrire une opération. Pour les articles expérimentaux identification des matériaux, des méthodes et d'une ou plusieurs techniques de caractérisation. Pour les articles théoriques, principalement identification de la méthode.
- 4) Création du champ « **champs personnalisés** ». A partir du dialogue établi entre le chercheur et J. M. Careil, une association de la valeur « Si » au champ « matériaux ». Définition d'une relation de synonymie entre les différentes écritures possibles du silicium (Si et Silicon)
- 5) Affectation des valeurs aux champs personnalisés (figure 6)



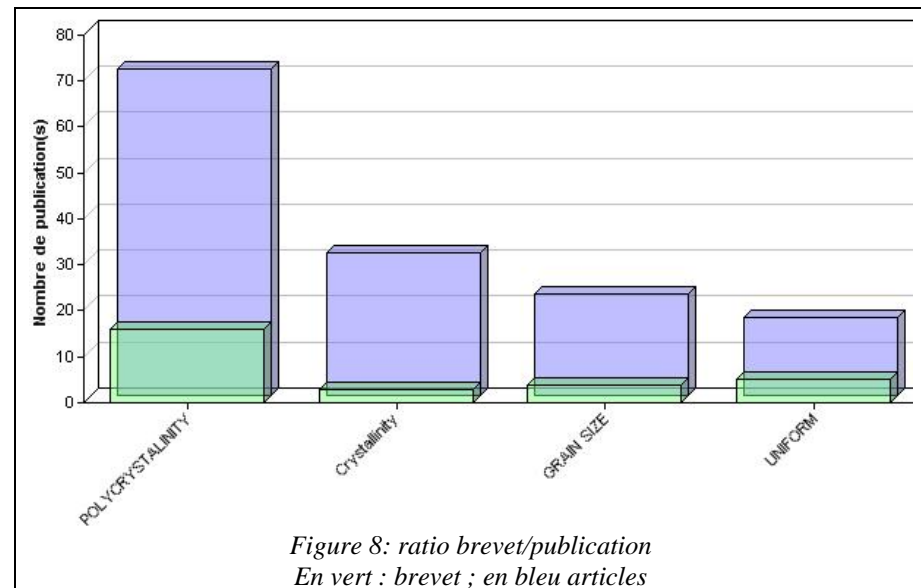
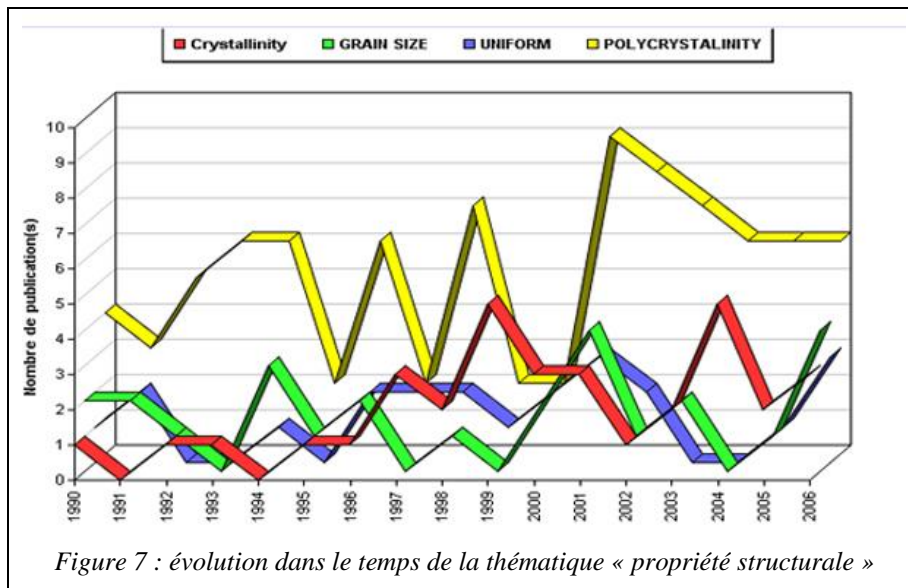
L'introduction des « champs personnalisés » donne de la plasticité à l'outil. Les champs reflètent l'état de connaissance du chercheur et traitent un large corpus de données suivant les thématiques désignées par le chercheur comme étant au cœur de ses problématiques de recherche.

Certes, il est aisé d'argumenter que le paramétrage de l'outil va orienter les résultats obtenus alors qu'il est notoire que l'innovation se situe souvent aux frontières du champ d'expertise. Cependant d'une part, les chercheurs académiques sont très souvent acteurs de discipline, et d'autre part, l'intérêt de cette approche réside dans les notices bibliographiques non affectées. Ces notices orphelines, n'ont pu être automatiquement affectées suivant les champs suivants, définis par le chercheur : « opération », « propriété_structurale », « méthode_caracterisation », « Propriété_optoelectronique », « disposition_fabrication », « dispositif_mesure ».

Deux hypothèses sont envisagées : Premièrement, il est judicieux de créer soit des champs personnalisés supplémentaires, soit d'élargir les valeurs possibles affectées aux champs personnalisés. D'autre part, ces notices restantes sont intéressantes car elles ne contiennent aucun champ précédemment défini et de fait se situent aux marges des connaissances du chercheur. Ainsi, nous avons trouvé un article relatif à la photographie qui s'appuie sur les procédés utilisés en photovoltaïque.

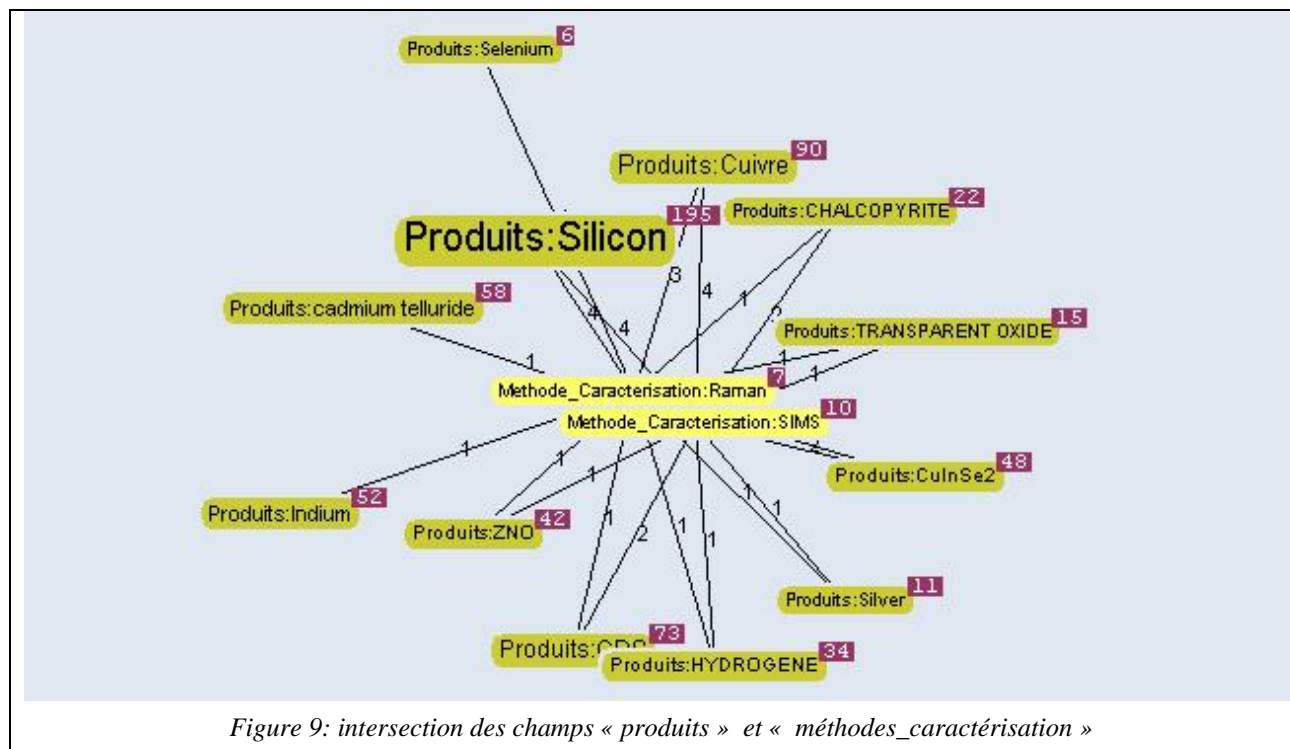
Concernant cette phase de personnalisation, signalons qu'il n'est pas toujours facile de définir à quel concept associer un terme : par exemple «Doping» signifie à la fois une propriété et une action.

A partir de chacun de ces concepts typés, il est possible de poursuivre l'analyse. Analyse rétrospective des différentes propriétés structurales. Ainsi la propriété structurale de polycrystallinity (figure 7) qui reste très majoritaire dans le corpus, a connu un pic de publications en 2002, avant de se stabiliser depuis 2005.



De la même façon il est judicieux de comparer le ratio publications/brevets des différentes techniques (figure 8).

Enfin, à partir de ces concepts typés (champs personnalisés) il est possible de les croiser en eux pour déduire des associations nouvelles, par exemple des matériaux et des méthodes (figure 9). Ce type d'association n'est pas forcément naturel pour le chercheur, c'est une autre lecture du corpus qui lui est proposé et qui peut suggérer des conditions expérimentales différentes, ou bien la mise en corrélation de conditions expérimentales pour optimiser la production d'un produit. L'outil de veille offre la possibilité de construire de nouveaux scénarii de recherche.



3.2 Positionnement du documentaliste/spécialiste de l'information

L'expérimentation menée nous conduit à faire le parallèle avec les profondes mutations engendrées par l'irruption de l'information électronique, directement consultable sur le poste de travail du chercheur. C'est le libre accès qui a fait le succès des périodiques électroniques et plus encore des bases de données de type « Chemical Abstracts » ou « Web of Science » auprès de la communauté scientifique. Fort de ce constat, comment est-il possible de favoriser une utilisation directe des outils de veille par les chercheurs, garant du succès de l'outil ? D'autre part, si la priorité d'un documentaliste est la qualité de l'information collectée, quel comportement/positionnement, un bibliothécaire/spécialiste de l'information doit-il adopter pour contribuer à une exploitation de qualité des outils de veille ?

Dans un premier temps S. Savina était convaincue qu'un tel outil ne pouvait être accepté par le chercheur qu'à la condition d'être en accès libre, en autonomie complète, sur son poste de travail, s'appuyant pour cela sur le succès des périodiques électroniques dans les pratiques de recherche d'information des chercheurs, avec une moyenne de 5000 articles téléchargés par mois depuis le site de « l'American Chemical Society » pour environ 300 chercheurs.

En observant le temps consacré par J M Careil à interroger le chercheur sur les notices feuilletées, pour identifier les champs majeurs et les valeurs attribuées, ainsi que les différentes expressions possibles pour formaliser ce champ, il semble que le documentaliste peut jouer un rôle dans ce processus

Il peut se positionner comme **médiateur** entre l'outil purement technique et la formalisation des connaissances du chercheur, et cela dans le but d'optimiser la qualité des résultats obtenus. Il suggère au chercheur une démarche pour formaliser les concepts à extraire et s'assure de l'intégration de ces concepts dans le système.

Cette attitude se rapproche d'une **démarche de conseil et d'accompagnement** et c'est peut être là que se situe une évolution du métier de documentaliste/professionnel de l'information, à savoir poursuivre l'accompagnement des chercheurs, comme ils ont su le faire lors de la décentralisation de l'accès à l'information par les ressources électroniques.

4 Conclusion

L'étude menée ici accrédite l'idée d'un type d'organisation dual entre d'une part, un centre de documentation, qui négocie les accès aux sources d'information, et intègre des professionnels qualifiés à l'utilisation d'outils de veille au cœur des équipes de recherche, et d'autre part des réseaux d'experts (chercheurs) seuls qualifiés dans l'étude de l'information collectée.

A l'heure où est publié cet article, nous sommes conscient que ces résultats ne sont que partiels. En effet, cette étude soulève plusieurs questions qui pourraient faire l'objet d'étude plus approfondie :

A - Quelle est la différence entre les résultats obtenus avec l'indexation de la base de données « Chemical Abstracts » et ceux obtenus selon l'approche décrite plus haut ?

Cette étude apporte un début de réponse. En général, les thésaurus utilisés par les bases de données sont gérées par des documentalistes dans une problématique de **recherche** d'information, ici dans l'approche proposée, nous travaillons à partir des connaissances du chercheur dans une problématique d'**analyse** d'information.

B - comment capitaliser le temps du paramétrage de l'outil, par le chercheur avec le soutien technique du documentaliste/professionnel de l'information ?

Plusieurs approches existent actuellement sur le marché. Les sociétés, TEMIS, fournissent avec l'outil d'extraction d'information, des cartouches de connaissances (skill cartridge) [7] ; INTELLIXIR [3] permet de paramétrer sur mesure son logiciel en fonction des connaissances du chercheur ; OPEN-MEVA [8], développé par C. Krumeich, s'appuie sur les travaux des sciences cognitives et de traduction de l'école D. Seleskovitch et part du postulat que la connaissance implicite d'un chercheur est unique, et que par conséquent chaque plan de veille est unique.

Il serait intéressant d'élargir le champ de cette expérimentation. De la mener par exemple, avec un panel de chercheurs du même laboratoire, pour mettre en évidence une unicité ou non du capital immatériel de ce même laboratoire ? L'enjeu de la capitalisation de la connaissance est stratégique pour la qualité d'un outil de veille.

5 Bibliographie

- [1] DOUSSET B., KAROUACH S., *manipulation de graphes de grande taille pour l'étude des réseaux d'acteurs et de réseaux sémantiques*, Acte du colloque de l'île Rousse, 2005
- [2] <http://outils.veille.inist.fr#outils>
- [3] <http://intellixir.com>
- [4] COURTIAL J.P., *Analyse dynamique des représentations sociales des chercheurs : le cas de l'autisme*, Papers on social representations 8, 1999, p.3.1-3.22
- [5] ROSTAING H., GIMENEZ TOLEDO E., KISTER J., *Système d'information pour l'aide à la gestion stratégique de la recherche dans un établissement public de recherche*, Actes du colloque VSST, 2004
- [6] BEGUIN A., *usage d'un thesaurus et développement de la pensée*, interCDI, n° 158, mars/avril 1999
- [7] GRIVEL L., GUILLEMIN-LANNE S., LAUTIER C., MARI A., *La construction de composants de connaissance pour l'extraction et le filtrage de l'information sur les réseaux*, ISKO International Society for Knowledge Organization, 5-6 juillet 2001, Université Paris X Nanterre, ISBN 2-9516737-0-1, pp 197-208, 2001
- [8] SCHARBOCHI E. KRUMEISH C. MALLET V, LAURENT D., *L'approche mémorielle, une méthode pour la veille documentaire scientifique et technique. Application dans le domaine de la recherche en médecine régénérative*, Actes conférence île Rousse, 2005