

DONNER DU SENS AUX BREVETS

Les méthodes de recherche

Fabienne MONFORT-WINDELS

Responsable Information et veille, Sirris

▪ Cet article constitue le 4^e article d'une série sur les brevets. Les deux premiers sont parus dans le numéro 2007/3 et le troisième dans le numéro 2008/1. L'information "brevets" peut être analysée par des méthodes qualitatives ou quantitatives. La recherche documentaire consiste à repérer dans la masse de documents quelques brevets intéressants. Un traitement statistique simple appliqué à un corpus permet d'extraire d'autres types d'information. Aujourd'hui, les données des brevets, jusqu'ici non codées, "muettes" pour les statisticiens, prennent du sens grâce aux nouvelles méthodes telles que le text mining. Ces techniques reposent sur des algorithmes qui repèrent des corrélations non immédiatement perceptibles entre les termes dans de grandes masses de documents. Elles permettent de détecter des "signaux faibles" à partir d'une information surabondante. Des outils intelligents vont encore plus loin. TRIZ, par exemple, est une méthode de créativité basée sur l'analyse d'un grand nombre de brevets.

▪ Dit artikel is het vierde in een reeks die verschillende aspecten van brevetten behandelt. Het vorige vindt u in het eerste nummer van dit jaar en de eerste twee in 2007/3. Voor de analyse van octrooi-informatie kunnen kwalitatieve of kwantitatieve methodes worden gehanteerd. De eerste zijn gericht op het vinden van enkele interessante octrooien (documentair opzoeken of deskresearch), terwijl de tweede gebaseerd zijn op een geautomatiseerde informatie-extractie en een eenvoudige statistische verwerking (bibliometrie) of een geavanceerde statistische verwerking (text mining). Data die tot nog toe niet gecodeerd waren, en voor de statistici dus 'nietszeggend' waren, krijgen nu wel betekenis dank zij nieuwe onderzoeksmethodes als text mining. Deze technieken steunen op algoritmes die niet onmiddellijk waarneembare verbanden vinden tussen de termen in grote massa's documenten. Ze kunnen "zwakke signalen" opsporen bij een overvloed aan informatie. De TRIZ techniek daarentegen is een creativiteitsmethode gebaseerd op de analyse van een groot aantal octrooien.

L'information "brevets" peut être analysée par des méthodes qualitatives ou quantitatives. Les premières visent à repérer quelques brevets

intéressants (recherche documentaire). Les secondes reposent sur une recherche automatisée et un traitement statistique simple (bibliométrie)

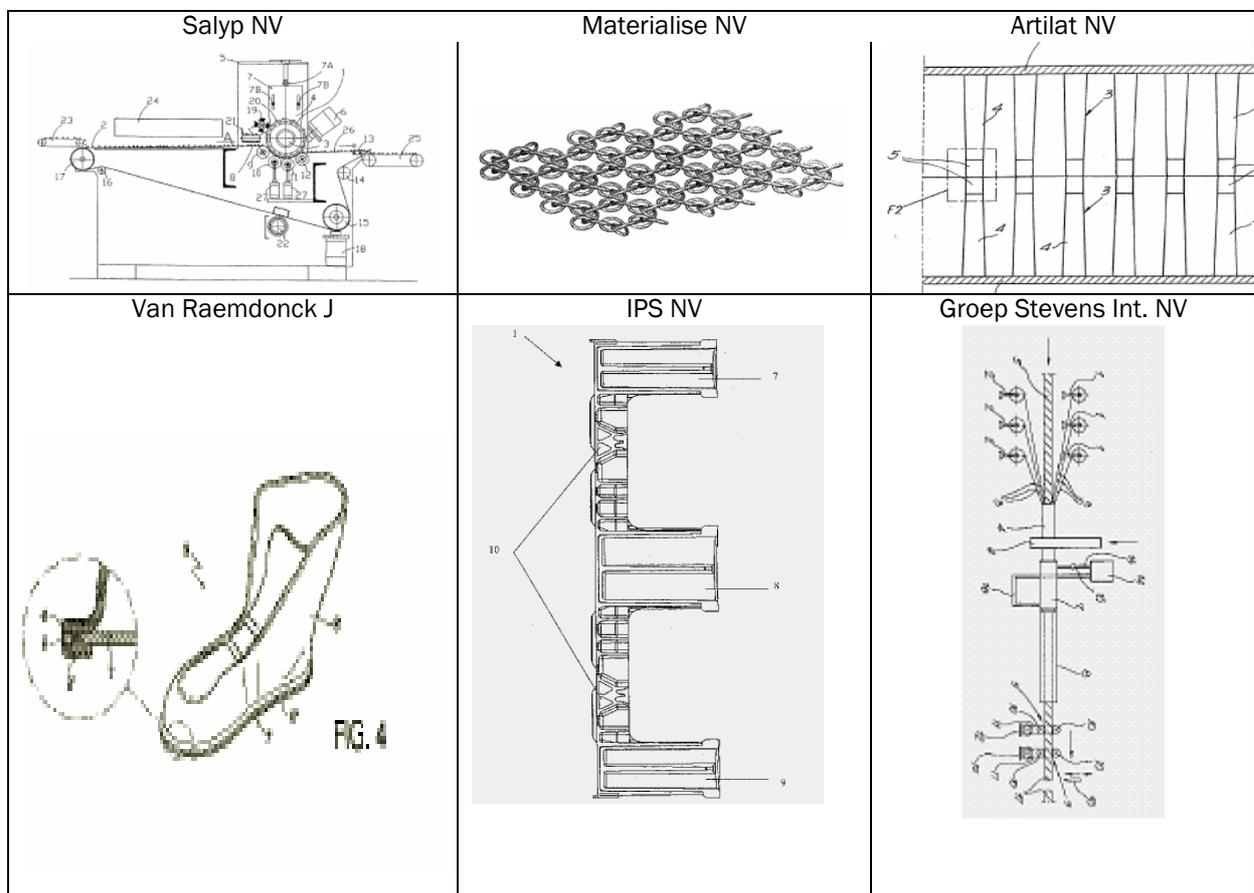


Fig. 1 : Brevets publiés en Belgique dans le domaine des plastiques au cours de 6 mois.

ou sophistiqué (text mining).

De nouvelles méthodes d'analyse qui apparaissent aujourd'hui ambitionnent de donner plus de sens aux brevets.

Ces méthodes ont dans la pratique des degrés de diffusion différents, correspondant au degré de maturité plus ou moins grand des outils informatiques.

La recherche documentaire (heuristique)

La recherche documentaire a pour objet de repérer dans les millions de documents les quelques brevets pertinents, à propos d'une question précisée par des noms de déposants, des mots-clés, des codes de classification... (exemple 17). (Un prochain article de cette série illustrera une recherche-type).

Exemple 17 : Il est ainsi possible de réaliser une recherche sur les brevets

- publiés en Belgique
- par des entreprises belges
- dans les 6 derniers mois
- et classés sous les rubriques B60 pour les véhicules, B21-B24 pour le travail du métal et B29 pour le travail des matières plastiques.

La recherche documentaire répond aux questions de base : Quels sont les développements dans tel domaine ? Qui les mène ?

Ce type de méthode représente aujourd'hui la quasi-totalité des recherches dans les brevets. Toutes les bases de données disposent de moteurs de recherche et offrent ces possibilités, avec des facilités plus ou moins grandes.

La recherche bibliométrique

La recherche bibliométrique est quantitative et se base sur des comptages (déposants, mots clés, classifications...) statistiques effectués sur un corpus de brevets, lui-même constitué des documents intéressants extraits par une recherche documentaire classique (exemple 18)

Exemple 18 : (Logiciel Aureka) : Une étude sur l'innovation dans les films d'emballage étanches

aux gaz, films dits à effet barrière, réalisée chez SIRRIS, donne le comptage par déposant, représenté sur la fig. 2.

Les informations brevets peuvent être utilisées dans la construction d'indicateurs statistiques et le comptage des données permet de structurer

| Assignee | Doc Count | Percentage |
|--|------------|------------|
| DAINIPPON PRINTING CO LTD | 70 | 13.1% |
| TOPPAN PRINTING CO LTD | 55 | 10.3% |
| NONE | 37 | 6.9% |
| CRYOVAC, INC. | 35 | 6.6% |
| TOYOBO CO LTD | 17 | 3.2% |
| MITSUBISHI POLYESTER FILM GMBH | 15 | 2.8% |
| UNITIKA LTD | 13 | 2.4% |
| HONEYWELL INTERNATIONAL INC. | 9 | 1.7% |
| MITSUBISHI GAS CHEM CO INC | 8 | 1.5% |
| TORAY IND INC | 8 | 1.5% |
| Number of assignments in Top 10 assignees | 267 | |
| Total number of assignments | 534 | |
| Number of documents after filter | 513 | |
| Total number of documents in group | 513 | |

Fig. 2 : Comptage par déposant lors d'une étude sur l'innovation dans les films à effet barrière pour le packaging.

un domaine de recherche en mesurant l'innovation par inventeur, laboratoire, entreprise, pays...

Les champs sur lesquels porte le comptage peuvent être de plusieurs types :

- caractéristiques techniques de l'invention : codes de classification, citations...
- auteurs de l'invention : inventeurs, déposants, adresses de ceux-ci...
- historique du brevet : dates de priorité, de la demande dans le pays concerné, de publication, de délivrance, pays dans lesquels la protection a été demandée...

Les questions qui trouvent réponse sont du type : Quels sont les pays qui déposent le plus dans tel domaine (exemple 19) ? Quelles sont les entreprises les plus actives en R&D (exemple 20) ? Quelles sont les applications les plus étudiées ?

Aujourd'hui, les recherches bibliométriques deviennent de plus en plus fréquentes vu la facilité de constitution du corpus et la généralisation des outils statistiques.

Exemple 19 : (Logiciel Tetralogie) : Les pays où on a déposé le plus de brevets dans le domaine de la métallisation des plastiques sont l'Allemagne, les États-Unis, puis le Japon et la France (dans les années couvertes par l'étude de SIRRIS).

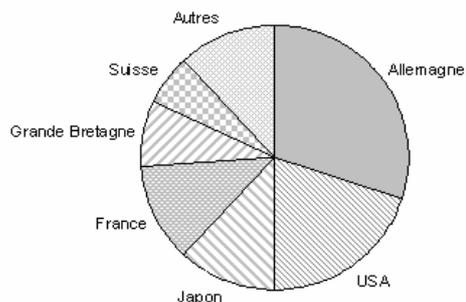


Fig. 3 : Pourcentage de brevets sur la métallisation des plastiques.

Exemple 20 : (Logiciel Matheo) : Dans le cadre de l'acquisition de ses équipements en thixomoulage, SIRRIS a évalué les perspectives de cette technologie. La figure 4 indique que les principales firmes innovantes dans la fonderie sous pression sont japonaises (B22d17 - Casting in which the metal is forced into a mould under high pressure) et qu'il s'agit surtout de motoristes.

En bibliométrie, on effectue souvent une analyse chronologique (exemple 21), pour mettre en évidence les évolutions des dépôts dans le temps : émergence de nouvelles technologies, nouvelles entreprises...

Exemple 21 : (Logiciel Micropatent) : L'évolution du nombre de brevets déposés dans le domaine des piles à combustible (H01M8) témoigne d'une importante augmentation des efforts de recherche à l'échelle internationale ces dernières années. En effet, des laboratoires de recherche mais aussi de grandes entreprises consentent des efforts considérables pour rendre les piles à combustible viables sur les plans technique et économique et préparer leur commercialisation (voir figure 5).

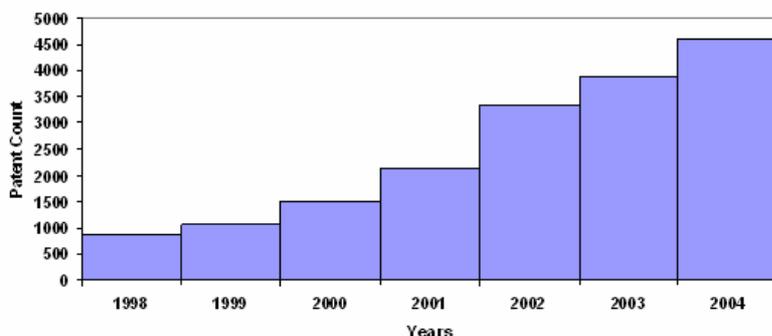


Fig. 5 : Nombre de brevets dans les piles à combustible en fonction du temps.

Une classification internationale spécifique, H01M8, porte sur les procédés ou moyens pour la conversion directe de l'énergie chimique en énergie électrique et plus particulièrement sur les éléments à combustible. Depuis 1998, 17.700 brevets ont été déposés dans cette catégorie.

L'analyse statistique des données de brevets est également utile sur un plan "collectif", par exemple pour orienter les efforts de R&D publics ou pour positionner une industrie.

L'analyse peut être plus qu'un simple comptage. Ainsi, par exemple, si l'on considère que les brevets sont de bons indicateurs de résultats en matière de R&D, on peut penser a priori qu'une entreprise qui dépose plus qu'une autre est plus active en innovation. Mais tous les brevets n'ont pas la même valeur. D'autres critères peuvent être pris en compte. On peut comptabiliser, par

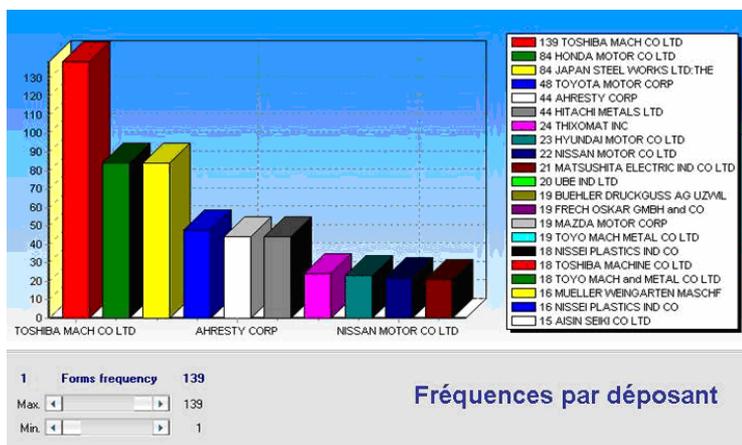


Fig. 4 : Les dix plus grands déposants de brevets dans la technique de la fonderie sous pression.

exemple, les brevets citants et les brevets cités, le nombre de revendications ou la taille des familles de brevets. Un brevet plus fréquemment cité que d'autres de la même époque est considéré comme de meilleure qualité ou ayant un plus grand impact (exemple 22) L'évaluation des liens entre brevets découverts par l'analyse des citations permet de cerner de plus près la R&D dans un secteur.

Exemple 22 : (Logiciel Aureka) : Le brevet WO0074763 de Georgia Tech Research Corp. publié en 2000 et qui traite de dispositifs de microaiguilles pour délivrer des médicaments ou prélever des fluides à travers les tissus biologiques de manière non invasive a inspiré 42 autres brevets (voir figure 6).

Le text mining

Les données jusqu'ici non codées, inutilisables pour les statisticiens, prennent du sens, et notamment le descriptif de l'invention lui-même. Le text mining repose sur des algorithmes qui repèrent des corrélations non immédiatement perceptibles entre les termes dans de grandes masses de documents (exemple 23). Il permet de détecter des "signaux faibles" à partir d'une information surabondante.

Exemple 23 : La figure 7 illustre une recherche à propos des brevets sur le thixomoulage (mise en forme à l'état semi-solide) du magnésium. L'histogramme représente la fréquence d'apparition de deux termes choisis – ici "alloy" et "thixotropic" – avec tous les autres termes importants du corpus de brevets examinés. Les fréquences relatives donnent une idée de la proximité des deux termes l'un par rapport à l'autre. On constate par exemple que si le terme "magnesium" est souvent associé au terme "alloy", ce qui est logique, le terme "thixotropic" apparaît rarement dans le même voisinage.

L'information pertinente est extraite du corpus préalablement constitué et est affinée par des outils d'analyse statistique lexicale ou sémantique. Les fonctions statistiques sont des calculs de fréquence univariée ou multivariée, des analyses en composantes principales, des analyses factorielles des correspondances, des classifications par partitions...

L'objectif est ici de mettre en évidence les technologies de substitution, les technologies associées, les technologies émergentes, les laboratoires, etc.

Les résultats sont présentés sous forme de visualisations graphiques : histogrammes, arbres de classification, cartes factorielles, réseaux de liens... (exemple 24).

Show HTML Report filtering options

Basic Report: Forward Citation Frequency For Most Cited Documents

source document list: **barriere 2000 F-List (DeDuped)** [Print-Friendly](#)

| Document ID | Assignee | Cited by | Percentage |
|--|--------------------------------------|------------|------------|
| WO0074763 | GEORGIA TECH RESEARCH CORPORATION | 42 | 13.3% |
| US6083628 | SIGMA LABORATORIES OF ARIZONA, INC. | 29 | 9.2% |
| EP0756931 | KUREHA KAGAKU KOGYO KABUSHIKI KAISHA | 18 | 5.7% |
| EP0707957 | CRYOVAC, INC. | 15 | 4.8% |
| WO0026973 | DELTA V TECHNOLOGIES, INC. | 15 | 4.8% |
| EP0721899 | CRYOVAC, INC. | 10 | 3.2% |
| EP0974452 | KUREHA KAGAKU KOGYO KABUSHIKI KAISHA | 10 | 3.2% |
| US6056141 | SAFTA S.P.A. | 10 | 3.2% |
| US6054212 | MITSUBISHI POLYESTER FILM GMBH | 9 | 2.9% |
| US6149995 | MITSUBISHI POLYESTER FILM GMBH | 9 | 2.9% |
| Number of citations in Top 10 documents | | 167 | |
| Total number of citations | | 315 | |
| Number of documents after filter | | 513 | |
| Total number of documents in group | | 513 | |

Fig. 6 : Classement des brevets les plus souvent cités.

Exemple 24 : (Logiciel Wordmapper) : Le diagramme de la figure 8 toujours dans le domaine du thixomoulage, est une présentation en cluster : chaque "étiquette" est un groupe de mots dont on étudie les relations avec d'autres. L'intensité des couleurs est fonction de la fréquence de ces groupes. On voit apparaître ici deux champs lexicaux distincts.

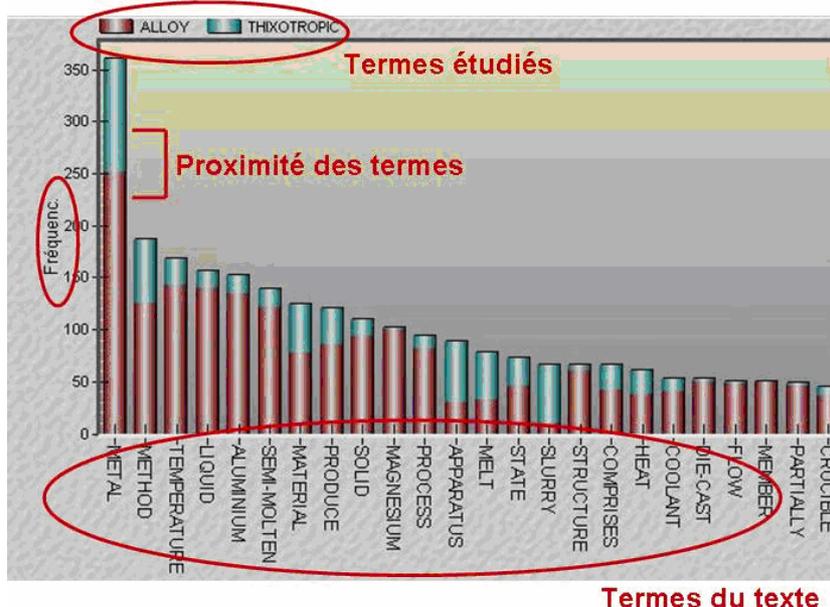
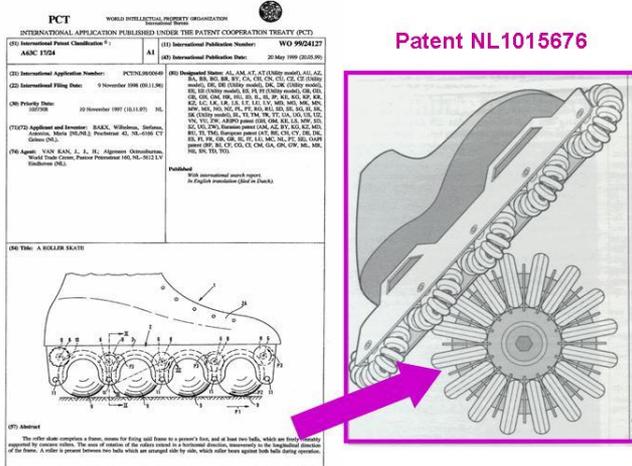


Fig. 7 : Fréquence d'apparition de deux termes choisis par rapport aux autres termes dans un corpus de documents.

vets concurrents (patent breaking) (exemple 27).
Il s'agit de réaliser une analyse fonctionnelle de la solution spécifique décrite dans un brevet, de remonter à partir de là à un modèle générique, puis de trouver d'autres solutions spécifiques qui remplissent les mêmes fonctions autrement.

Exemple 27 : L'exemple illustré sur la figure 12 a été présenté au SIRRIS lors d'une conférence sur le patent breaking par ITER, bureau spécialisé dans l'application de la méthode TRIZ : il concerne l'amélioration de rollers à partir de l'examen des brevets existants et a abouti à un nouveau brevet (NL1015676), innovant par rapport à l'état de l'art.



Fabienne Monfort-Windels
Sirris
Rue du Bois Saint-Jean, 12
4102 OUGRÉE
fabienne.windels@sirris.be
Mai 2008

Fig. 12 : Un concept innovant de rollers à partir de l'étude des brevets existants.