



La valeur ajoutée de Derwent

François Libmann

Derwent a publié récemment dans *STNews* (avril 2010) un article expliquant que la correction des erreurs détectées dans les documents de brevets était une partie importante de sa valeur ajoutée (les autres étant la réécriture des titres, l'écriture d'abstracts structurés et l'ajout d'une indexation sophistiquée).

En 2009, plus de 3,13 millions de brevets ont été chargés dans DWPI, une augmentation de 14,5 % par rapport à l'année précédente.

Les documents entrants sont analysés automatiquement et mis dans le format Derwent. Ce programme permet aussi de déceler l'essentiel des erreurs présentes dans les documents, telles que numéro de priorité incorrect, erreur dans le nom du déposant, absence de code CIB. Ces brevets sont alors corrigés à la main ce qui, on l'imagine, est relativement coûteux.

A titre d'exemple, pour les deux premiers mois de l'année, 6 198 documents avaient des données de priorité absentes ou incorrectes et des codes CIB ont été ajoutés manuellement sur 450 documents PCT, qui en étaient dépourvus.

Du côté de Questel et, plus précisément, de FamPat, différents algorithmes sont utilisés pour vérifier les données

de brevets, mais Questel ne fait pas d'intervention manuelle ne souhaitant pas, du moins pour l'instant, ajouter une fonction d'éditeur à son rôle de producteur.

Les erreurs ou invraisemblances détectées, notamment pour les priorités, sont signalées par Questel à l'OEB. Par ailleurs, un champ non accessible aux utilisateurs permet d'avoir une traçabilité pour tout ce qui concerne les changements relatifs à une priorité.

Lors de la récente réunion du CFIB (Club francophone de l'information brevet), un participant a signalé que pour les PCT non examinés, il était assez fréquent qu'aucun code CIB ne figure dans la référence du document.

Une recherche dans PlusPat de Questel a effectivement fait apparaître, en particulier, des brevets PCT récents d'origine US non examinés (catégorie A2) sans code CIB, lequel code est en général ajouté après la réalisation du rapport de recherche (catégorie A3).

Curieusement, l'absence de code CIB n'empêche pas toujours la présence de codes ECLA. Si l'on fait le rapport entre le nombre de références sans code CIB dans les brevets PCT et le total des brevets PCT dans PlusPat, on arrive à 0,2 %.

Dans Espacenet, sur les 100 brevets PCT les plus récents, 2 n'ont pas de code CIB, soit 2 %. Cela est à la fois beaucoup et peu et illustre une fois de plus le fait que la recherche brevet n'est pas une science parfaitement exacte.

Pour en revenir à l'article de *STNews*, celui-ci mentionnait à titre d'exemple la demande de brevet US 20100059621 (A), dans laquelle le nom du déposant était devenu AIRBUS Operations. Le brevet est sous priorité française, le nom du déposant étant correctement orthographié dans le brevet français.

La correction a permis dans WPI de rattacher le brevet US au brevet français et, surtout, de lui donner le code déposant EADS, puisqu'AIRBUS en est une filiale.

Nous avons regardé à différents endroits si l'on retrouvait ou non cette faute :

■ **dans FamPat**, un seul document regroupe le brevet US et le brevet français. On trouve mentionnés trois *patent assignees* : AIRBUS France, AIRBUS Operations et le fameux AIRBUS Operations. Curieusement, l'original *Applicant Assignee* est AIRBUS Operations, alors que le brevet français, dans lequel il n'y a pas de faute, est antérieur ;

■ **dans Inpadoc/Fam & Legal Stat** (fichier n° 345 sur Dialog), on trouve aussi les deux documents

dans la même famille, avec la faute dans le brevet US et l'on peut supposer que c'est de même nature pour les bases équivalentes sur STN.

Cela illustre à l'évidence, ce que l'on savait déjà, à savoir que les familles sont définies par les priorités, ici dans un cas particulièrement simple ;

■ **dans World Patents Index**, le document synthétique comporte le nom correct, mais dans les "documents originaux" qui lui sont rattachés, le brevet américain comporte la faute ;

■ **dans Espacenet**, on trouve le brevet US déposé par AIRBUS, mais il est relié au brevet français AIRBUS ;

■ **dans les banques de données ne recensant que les brevets américains**, comme USAPPS sur Questel, Claims US Patents ou USPATFull sur Dialog, on ne retrouve que le nom AIRBUS et le numéro (seulement) du brevet français.

On voit donc l'intérêt qu'il y a, lorsque l'on recherche par nom de déposant, à interroger des banques de données structurées en famille, plutôt que des banques de données nationales, pour parer à ces erreurs de saisie ; nous ne parlerons pas ici des ambiguïtés sur les transcriptions du nom des inventeurs chinois. C'est un sujet en soi !.