

# La société infoapps au salon ICIC

FRANÇOIS LIBMANN

## ChemAnalyser (infoapps) avec ontochem IT Solution

**ChemAnalyser** de la société allemande **infoapps** développé en collaboration avec **OntoChem IT Solution**, une autre société allemande, est présentée comme la plus grande et la plus complète des banques de données dans le domaine de la chimie.

Le contenu est constitué d'une part d'informations issues de 105 offices de brevets parmi lesquels 37 offices proposent le texte intégral en anglais, ou à défaut en traduction machine avec le plein texte original, le tout sans restriction par des classes, contrairement à Patentscope.

Cela représente environ 100 millions de documents.

Par ailleurs, 80 millions de documents sont des articles scientifiques (abstracts ou textes intégraux selon les cas), des livres ou des comptes-rendus de conférences. La plupart de ces documents sont référencés dans Scopus, Embase, Medline/PubMed ou ScienceDirect. Mais tous les utilisateurs n'ont pas accès à l'ensemble de ces 80 millions de documents. En effet, pour les services payants comme Scopus, il est nécessaire d'avoir signé un contrat avec ce service pour accéder aux documents qu'ils proposent.

On y trouve enfin des informations d'autres sources comme ChEMBL, Clinical Trials et d'autres sites Web tels que ceux de la FDA, l'EMA et l'EFSA. Il est possible en outre d'inclure dans le corpus des documents internes.

Infoapps s'appuie sur la reconnaissance automatique considérant que, sauf quelques exceptions, elle est plus efficace que la reconnaissance humaine. Cette reconnaissance se fait sur les structures ou sur les termes, leurs relations et leurs rôles dans différents domaines. Ces termes sont classés et peuvent être enrichis de différentes façons (synonymes, variations orthographiques ou linguistiques).

Plus de 6 milliards (!) de composés chimiques et substances ont été reconnus dans ChemAnalyser de même que 76 millions de structures chimiques uniques définies par leur numéro InChi.

La recherche est baptisée «recherche cognitive». Le système est réactif en proposant par exemple de compléter des termes ou des structures, d'inclure des synonymes ou des traductions de termes et de choisir un domaine en cas d'ambiguïté d'un terme, par exemple le terme grain. Un gros travail d'inventaire dans des domaines comme les plantes comestibles a été réalisé depuis une dizaine d'années. La banque de données est actualisée de façon

hebdomadaire et l'ensemble du backfile est réindexé tous les trimestres.

Infoapps a, par ailleurs, une offre conséquente dans le domaine des brevets, (voir l'article intitulé «*sem-ip.com* : l'offre brevet d'infoapps» dans ce même numéro).

## sem-ip.com : l'offre brevet d'infoapps

Outre son offre en chimie, **infoapps** propose également une offre significative dans le domaine brevets.

Sa banque de données *sem-ip.com* permet de rechercher dans une centaine de millions de documents regroupés en 50 millions de familles Inpadoc. Ces documents proviennent de 105 offices et, pour 37 d'entre eux, le texte intégral en anglais natif ou une traduction machine est disponible, de même que le texte intégral original.

Il est possible d'effectuer des recherches, soit avec l'interface dédiée à la recherche booléenne, soit en entrant un texte ou des termes dans l'interface de recherche sémantique. Des outils classiques d'analyse poussée des résultats sont disponibles.

Les résultats des veilles et éventuellement des recherches effectuées dans *sem-ip.com* sont versés dans **infoPatent** qui est une banque de données in-house, préservant ainsi la

confidentialité, et constituant, éventuellement avec d'autres documents, qui ne seront pas nécessairement des brevets, une base de connaissance pour l'entreprise.

Cet outil permet de diffuser de façon ciblée les résultats des différentes veilles aux ingénieurs concernés en leur proposant différentes options pour réagir.

On notera que les mêmes possibilités de recherche booléenne ou sémantique que dans *sem-ip.com* sont disponibles dans **infoPatent**.

Il existe enfin une option baptisée *infoPatent-classifier* qui est un outil de classification automatique utile, par exemple, lorsqu'une veille génère un grand nombre de documents.

Pour limiter le nombre de documents, s'ils sont trop nombreux, l'utilisateur donne des éléments des filtres qu'il souhaite utiliser pour la sélection des brevets issus de la veille. Un système d'apprentissage est enclenché pour optimiser la sélection. Par la suite, l'utilisateur définit le degré de similarité requis pour l'inclusion ou non des documents dans les structures qu'il a préalablement définies.

*infoapps* fait état de 60 000 utilisateurs dans le monde.

L'un des commerciaux d'*infoapps* est français, et peut être contacté par email à l'adresse [r.marche@infoapps.de](mailto:r.marche@infoapps.de).