

REVIEW: Competitive intelligence and patent analysis in drug discovery: Mining the competitive knowledge bases and patents

Grandjean, N., Charpiot, B., Pena, C.A., Peitsch. M.C (2005)

(Recensió per M. Welling Flensburg)
e-mail: m.welling.flensburg@gmail.com

L'article pretén establir la relació que existeix entre l'anàlisi de patents i la intel·ligència competitiva en el context de la indústria farmacèutica, en particular en el procés de desenvolupament de fàrmacs.

El text analitza com les tècniques d'anàlisi de patents i d'exploració de bases de dades s'apliquen i com aquests mètodes poden ajudar en la tasca d'aconseguir informació sobre les activitats de la competència, activitats de R+D que s'estan realitzant, camps amb potencial de creixement, o sobre les col·laboracions que s'estan fent en aquest camp. L'autor exposa els avantatges i els desavantatges de l'anàlisi de patents, encara que conclou que les patents segueixen essent la font més fiable d'informació sobre activitats de R+D.

En primer lloc, els autors evidencien el retard temporal que existeix entre el moment de la inscripció d'una innovació a les oficines de patents, i el moment en que es publiquen aquestes patents. Aquest retard, d'uns 18 mesos, representa un desfasament entre l'estat de la tècnica "real" i el que es publica, la qual cosa representa un inconvenient. Malgrat aquest problema, l'anàlisi de patents o patentometria és el millor recurs existent per conèixer l'estat de les activitats de R+D d'una indústria determinada. Donada la importància dels registres de patents, la seva qualitat determina en gran mesura la utilitat que se li pugui donar (especialment en els casos en els que es vulgui conèixer millor l'estratègia d'una empresa basant-se en les patents que hagi enregistrat). En aquest context, el principal problema dels arxius de patents primaris (aquells que recullen el que és publicat per les oficines de patents) radica en la seva complexitat general que dificulta una interpretació estratègica. En relació a les bases de dades comercials, indexades per tercers, el seu principal problema és la manca de detall i profunditat en els continguts.

A continuació, els autors expliquen com, davant de les dificultats per interpretar sistemàticament les bases de dades existents, pot ser interessant realitzar un procés d'extracció d'informació útil de les patents. El procés en qüestió pot ser relativament senzill, en els casos en els que calgui detectar i normalitzar conceptes de forma més o menys sistemàtica, o pot ser més complex. En aquest segon cas, l'extracció requereix mètodes més complexes de processament de text, que permetin obtenir un conjunt de dades susceptibles d'anàlisi, exploració i interpretació posterior. Un bon exemple d'extracció complexa seria el cas de conceptes com 'nom de l'empresa', "autors de la invenció" o "models d'actuació" que poden ser extrets i utilitzats per més d'una base de dades de segon nivell, en funció del criteri que regeixi cada base de dades en qüestió. D'aquesta manera, aquests conceptes poden utilitzar-se a mode de connectors entre aquestes bases de dades secundàries dins de l'arxiu principal.

Els principals passos del procés d'extracció d'informació de les patents són:

Nº de fase	Fase
1	<i>Zoning</i> : marcar frases rellevants identificades segons la seva importància científica.
2	<i>Parsing</i> : extracció de les frases a un arxiu de dades
3	Normalització: es fa aplicant regles (algoritmes) o bé assignant termes als sinònims
4	Extracció de conceptes: des del text, tenint en compte l'arxiu de dades creat en els passos 2 i 3, és a dir, que extreu conceptes considerant els sinònims que poden adoptar els conceptes i el context en el que es troben
5	Categorització: ordenació dels conceptes en sub-arxius
6	Aplicació d'un algoritme de taxonomia que permeti navegar per la base de dades en coherència amb el criteri de recerca que es desitgi
7	Relacionar conceptes: les relacions s'estableixen aplicant regles d'anàlisi sintàctic.

El procés anterior permet crear bases de dades amb funcionalitat i sensibilitat als criteris de recerca que siguin d'interès per a l'usuari. Aquestes bases de dades han d'estar actualitzades i hauran de ser científicament coherent, matemàticament vàlides, i susceptibles de poder ser explorades de forma funcional.

Per a explorar i analitzar les bases de dades, se solen utilitzar tècniques específiques d'exploració, que permetin establir relacions entre distintes variables concretes. A més se solen utilitzar també eines gràfiques que il·lustren millor les relacions entre variables i també són capaces de detectar nous patrons inèdits o altres informacions que abans no es coneixien.

Per últim, l'article esposa les alternatives comercials disponibles, tant productes per al desenvolupament d'anàlisi de patents i exploradors de bases de dades, com paquets comercials ja desenvolupats.