

La ricerca documentale dei brevetti

Problematiche e sviluppi futuri

Massimo Barbieri

Giuseppe Conti

Politecnico di Milano
Technology Transfer Office
massimo.barbieri@polimi.it
giuseppe.conti@polimi.it

Introduzione

Quando si effettua una ricerca sullo stato dell'arte brevettuale, la prima domanda che ci si deve porre è la seguente: quale database devo scegliere affinché la ricerca da svolgere sia la più completa possibile?

A volte, quando non si hanno soldi a disposizione ma solo un computer e un collegamento a Internet, la soluzione è semplice: un database gratuito.

Ovviamente questo servizio non è da intendersi come scarso o poco affidabile, ma come uno strumento che è stato pensato per fornire al pubblico la possibilità di avvicinarsi alla ricerca documentale ed è stato volutamente limitato in alcune funzionalità per evitare la concorrenza con i produttori di database commerciali.

La principale differenza tra database gratuiti (per es. Esp@cenet, banca dati USPTO, JPO, Depatisnet) e a pagamento (per es. Dialog, Questel-Orbit, Delphion, Micropatent) consiste nella capacità di questi ultimi di fornire risultati attinenti con l'oggetto della ricerca in tempi minori, poiché l'informazione è più completa e dettagliata (i titoli del Derwent World Patents Index, per esempio, sono senz'altro più utili dei titoli originali dei brevetti, proprio perché scritti in funzione di una ricerca



documentale), nonché nella possibilità di eseguire rielaborazioni dei dati raccolti.[1]

Strategie di ricerca

Una ricerca sullo stato dell'arte brevettuale (RSA) è effettuata generalmente in funzione del destinatario della ricerca stessa (azienda, centro di ricerca, ente finanziatore ecc.), per motivi quali:

- indirizzare la ricerca di centri di ricerca ed enti finanziatori;
- conoscere lo stato dell'arte prima di iniziare una qualsiasi ricerca;
- evitare di depositare brevetti privi dei requisiti o in contraffazione;
- avere notizia di potenziali *competitors* o di eventuali licenziatari;
- monitorare la concorrenza, l'investimento in tecnologia e prodotti innovativi.

Una RSA serve principalmente per valutare i requisiti di novità e attività inventiva nella fase di valutazione preliminare di un'invenzione, ovvero nel periodo che precede la stesura della domanda di brevetto.

Un altro aspetto da considerare riguarda il monitoraggio del portafoglio brevetti di un'azienda, per individuarne l'evoluzione tecnologica, ma anche per accertare eventuali casi di contraffazione.

La RSA è utile anche nella fase di decisione strategica di seguire o meno una determinata linea di ricerca ed evitare di ottenere un dispositivo o un procedimento già brevettato, e quindi di dissipare tempo e denaro utile per acquisire conseguentemente una licenza che permetta di produrre, ovviamente non in regime di monopolio.

Metodi e procedure di ricerca

Le strategie per eseguire una ricerca brevettuale sono molteplici, ma ciascuna di esse possiede un denominatore comune, individuato nell'elaborazione di una serie di "concetti" in grado di definire un'invenzione.

I concetti possono essere espressi in molti modi e spesso i ricercatori si sforzano di trovare le parole e

i sinonimi appropriati per caratterizzare un'invenzione.[2]

La combinazione di queste parole chiave (attraverso i classici operatori booleani e di prossimità) può portare a individuare uno o più documenti attinenti, a partire dai quali è possibile perfezionare ulteriormente la ricerca.

Tuttavia, le ricerche per parole chiave sono inadeguate in molti settori della tecnologia, per esempio perché non esiste ancora una terminologia standard, o perché le invenzioni possono essere descritte in un numero quasi infinito di modi alternativi oppure perché l'uso dei sinonimi richiede la loro traduzione in differenti lingue.[3] Inoltre, informazioni specifiche possono essere assenti nel testo, ma presenti solo nei disegni.

Dal momento che i brevetti sono classificati, la ricerca di un determinato concetto non deve essere limitata a una serie di parole chiave, ma può utilizzare anche un simbolo di classificazione (IPC, ECLA ecc.).[4]

Una combinazione di concetti e di simboli di classificazione può condurre all'ottenimento di un numero di documenti attinenti superiore rispetto alla semplice ricerca per parole chiave.

Si tratta, in fondo, di un processo iterativo che termina quando il ricercatore esperto ha la sensazione che ulteriori sforzi non portano ad alcun risultato utile.

La classificazione internazionale dei brevetti

La classificazione internazionale dei brevetti (International Patent Classification, IPC) fu creata ufficialmente nel 1971, quando fu stipulato l'Accordo di Strasburgo, e divenne operativa il 7 ottobre 1975 in tredici nazioni.[5]

Tale sistema fu inizialmente ideato per facilitare le ricerche realizzate



sulle copie cartacee, come strumento per creare pile di documenti di dimensioni ragionevoli in modo da poterli visionare facilmente uno per uno.

Successivamente, con l'avvento delle tecnologie informatiche, la classificazione IPC ha acquisito una maggiore importanza come strumento di ricerca, conservando quel "valore universale" indicato quasi trent'anni fa nell'Accordo di Strasburgo.[6] Infatti, le ricerche con parole chiave dimostrano parecchie limitazioni, in parte dovute al linguaggio, ma anche per il problema dei sinonimi e di una terminologia che cambia frequentemente.

Tuttavia, così come è stato concepito e strutturato, il sistema di classificazione internazionale presenta alcuni problemi, *in primis* la mancanza di un aggiornamento continuo.

Attualmente è in uso la settima edizione, in vigore dal 1° gennaio 2000; la precedente edizione è stata valida nel periodo 1995-2000, un tempo effettivamente troppo lungo se confrontato con l'evoluzione della tecnologia.

La classificazione internazionale non è da intendersi come un sistema omogeneo, ma si tratta piuttosto di sette sistemi che hanno alcuni elementi in comune ma anche elementi nuovi. Inoltre,

l'IPC è essenzialmente un sistema di classificazione orientato alla funzione del trovato, piuttosto che verso l'applicazione; le varie categorie rispecchiano, pertanto, ciò che l'invenzione realizza anziché il settore industriale cui si applica.[7,8]

Il fatto che i brevetti non siano riclassificati quando entra in vigore una nuova edizione rende di fatto la classificazione internazionale virtualmente inutilizzabile per molti dei suoi scopi.[9] Per questo motivo alcuni maggiori uffici nazionali brevetti hanno elaborato sistemi di classificazione interni basati sull'IPC.[10]

La successiva edizione dell'IPC (e precisamente l'ottava), che sarà pubblicata nel giugno 2005 ed entrerà in vigore a partire dal 1° gennaio 2006, dovrebbe risolvere gran parte dei problemi summenzionati, diventando un utile strumento per le ricerche brevettuali.

Si tratterà di un sistema a due livelli: uno definito *core* e l'altro *advanced*. Il livello *core* sarà sottoposto a revisione ogni tre anni, mentre il livello *advanced*, che rappresenterà un'ulteriore elaborazione del livello *core*, includerà sia quest'ultimo sia altri sottogruppi, e sarà revisionato ogni tre mesi.[11]

Un ulteriore sforzo dovrà essere dedicato alla riclassificazione dei brevetti, nonché a rendere omogenee le varie classificazioni.

Per esempio, si supponga di dover effettuare una ricerca nel settore delle penicilline. Esistono due tipologie di classificazione: A61K 31/43 (preparazioni farmaceutiche contenenti penicillina) e C07D 499/00 (composti di penicillina, inclusi i metodi di sintesi). La ricerca in una sola delle due classificazioni non consentirebbe l'ottenimento di un'informazione completa.

La classificazione internazionale dei brevetti è strutturata in otto *sezioni* (lettera A-H), ciascuna delle quali è suddivisa in un certo nu-

mero di *classi* (il cui simbolo è una doppia cifra), a loro volta comprendenti una serie di *sotto-classi* (rappresentate da una lettera A-Z), di *gruppi* (rappresentati da numeri da una a tre cifre, mentre elettronicamente le cifre mancanti sono sostituite da zeri) e di *sotto-gruppi* (comprendenti una notazione numerica da due a cinque cifre), separati dai gruppi mediante una barra spaziatrice.[12]

Un esempio di simbolo di classificazione è il seguente: H04K 3/00.

Altri sistemi di classificazione

Gli altri sistemi di classificazione sono basati sulla stessa filosofia dell'IPC, ma si tratta di sistemi dinamici (ovvero aggiornati di frequente) e costituiti da una singola edizione.

Nel caso della classificazione europea (European Classification, ECLA), il cui sviluppo iniziò nei primi anni Settanta, sono stati creati più sottogruppi interni rispetto all'IPC, in modo da ridurre il numero di documenti da consultare.[13] Alla fine del 1997, per esempio, la classificazione ECLA conteneva circa 120.000 gruppi, rispetto ai 67.000 dell'IPC.[14] I documenti vengono inoltre riorganizzati per migliorare la consistenza e ridurre la sovrapposizione.

Un esempio dell'efficacia dei codici ECLA nella ricerca documentale è fornito dal settore della chimica combinatoriale. Per questa disciplina non c'è ancora una classe IPC (sarà proposta la classificazione C40B), mentre esistono diverse classi ECLA, tra cui B01J19/00C (dispositivi per la chimica combinatoriale), C07B61/00L (sintesi di librerie di composti chimici) ecc.[15]

In alcuni settori l'utilizzo di schemi di indicizzazione (per es. il sistema ICO, In Computer Only) fornisce ulteriori modalità per accedere alla

documentazione. I codici ICO sono principalmente usati per indicizzare aspetti secondari menzionati nei documenti brevettuali.[16] L'ufficio brevetti giapponese utilizza un sistema denominato File Index (FI), che è basato sulla quarta, quinta e sesta edizione dell'IPC e che comprende approssimativamente 185.000 sottodivisioni.

La struttura è molto simile a quella dell'IPC, con le stesse sezioni, classi, sottoclassi, gruppi e sottogruppi; ciò che caratterizza questa classificazione è la presenza di un simbolo di sottodivisione (costituito da un numero di tre cifre) ed eventualmente di un cosiddetto *file discrimination symbol* (rappresentato da una lettera). Per esempio, la classe FI "G06F 9/00, 101" corrisponde alla classe IPC "G06F 9/00".

Accanto al File Index è stato creato anche il sistema di classificazione denominato F-terms, pensato in funzione delle ricerche brevettuali; tuttavia, mentre la classificazione FI è applicata a tutti i documenti (sia domande di brevetto sia modelli d'utilità), gli F-terms sono utilizzati solo per i documenti di specifici settori della tecnologia.[17]

La revisione della classificazione nazionale degli Stati Uniti (USPC), creata nel 1836, avviene ogni due mesi e quindi anche questa classificazione è uno strumento molto utile per le ricerche brevettuali.

Tuttavia, rispetto all'IPC, le classi USPC sono molto più suscettibili di variazione nella fase di esame di merito e di opposizione con una variazione del 12%, rispetto al 4% stimato per gli IPC nelle domande di brevetto europeo.[18]

Conclusioni

La ricerca effettuata mediante parole chiave non è né efficiente né precisa; infatti, dà luogo a risultati incompleti e richiede molto tempo

per la selezione dei documenti pertinenti.

Il problema principale dell'accesso ai documenti brevettuali è la lingua. Le domande di brevetto europeo, per esempio, possono essere depositate in una delle tre lingue ufficiali dell'European Patent Office (inglese, francese e tedesco), ma solo le rivendicazioni sono tradotte in tutte e tre. Dal momento che esiste un limitato controllo su queste traduzioni, l'informazione critica può essere reperita solo nella lingua d'origine.

Per ottenere un'informazione completa sarebbe necessario, in teoria, tradurre la *query* di ricerca in tutte le lingue (ovviamente è indispensabile che la traduzione della terminologia tecnica sia corretta) e poi applicare le tecniche di ricerca standard in ciascuna di esse. Le difficoltà aumentano nei brevetti, per il fatto che solitamente vengono generati neologismi per descrivere nuove invenzioni.

Un'altra fonte d'errore nella ricerca full text è costituita dal fatto che i documenti brevettuali sono attualmente ancora depositati in forma cartacea; il brevetto deve essere scannerizzato e l'immagine inviata a un software OCR che trasforma l'immagine in testo. L'accuratezza di tali sistemi è solo del 95%; pertanto, nel caso di documenti di oltre cento pagine, l'errore può essere elevato. [19]

Un'altra limitazione consiste nel fatto che la pubblicazione di una domanda di brevetto avviene dopo diciotto mesi dalla data di deposito della stessa (o comunque dopo un periodo minimo di tre mesi) e in questo lasso di tempo non è possibile reperire informazioni.

Tutto ciò non significa che una ricerca tramite parole chiave sia inutile: in molti casi serve a trovare una "traccia", da cui partire per perfezionare la ricerca.

Ciò che è stato argomentato nei precedenti paragrafi serve ad av-

valorare il ruolo primario dell'informazione nella ricerca scientifica. Esistono, tuttavia, alcune limitazioni nel reperimento delle informazioni, limitazioni che dovranno essere progressivamente eliminate (per esempio con l'aggiornamento continuo della classificazione IPC) per poter operare in un regime di certezza.

Bibliografia

- [1] PAUL SCHWANDER, *An evaluation of patent searching resources: comparing the professional and free on-line databases*, "World Patent Information", 22 (2000), p. 147-165.
- [2] DAVID HULL, et al., *Language technologies and patent search and classification*, "World Patent Information", 23 (2001), p. 265-268.
- [3] ANDERS BRUUN, *Development of the IPC as a search tool*, "World Patent Information", 21 (1999), p. 97-100.
- [4] PAUL SCHWANDER, *Prior art searches: a must for innovative smes*, <http://www.wipo.int/sme/en/documents/prior_art.html>.
- [5] MICHELLE LYON, *Language related problems in the IPC and search systems using natural language*, "World Patent Information", 21 (1999), p. 89-95.
- [6] MIKHAIL MAKAROV, *The process of reforming the International Patent Classification*, "World Patent Information", 26 (2004), p. 137-141.
- [7] PASCAL FAUCOMPRÉ – LUC QUONIAM – HENRI DOU, *The function-application relation through a link between classification and indexing*, "World Patent Information", 19 (1997), 3, p. 167-174.
- [8] JOHN BRENNAN, *Selection of data from the mass of information*, "Library Acquisitions: Practise and Theory", 21 (1997), 3, p. 303-317.
- [9] STEPHEN KUNIN, *IPC in the new millennium – challenge and opportunity*, "World Patent Information", 21 (1999), p. 101-108.
- [10] JAMES CALVERT – MIKHAIL MAKAROV, *The reform of the IPC*, "World Patent Information", 23 (2001), p. 133-136.
- [11] MICHELLE LYON, *Language related problems...*, cit.
- [12] STEPHEN ADAMS, *Using the International Patent Classification in an online environment*, "World Patent Information", 22 (2000), p. 291-300.
- [13] JÜRGEN RAMPPELMANN, *Classification tools at the EPO*, "World Patent Information", 18 (1996), 3, p. 149-153.
- [14] ID., *Classification and the future of the IPC – the EPO view*, "World Patent Information", 21 (1999), p. 183-190.
- [15] NIELS STEVNSBORG, *Searching combinatorial chemistry: from biochips to catalysts and beyond – an EPO perspective*, "World Patent Information", 24 (2002), p. 13-23.
- [16] J. VAN THIELEN, *Online classification and indexation of documents at the European Patent Office*, "World Patent Information", 20 (1998), 1, p. 17-20.
- [17] IRENE SCHELLNER, *Japanese File Index classification and F-terms*, "World Patent Information", 24 (2002), p. 197-201.
- [18] STEPHEN ADAMS, *Comparing the IPC and the US classification systems for the patent searcher*, "World Patent Information", 23 (2001), p. 15-23.
- [19] DAVID HULL, et al., *Language technologies and patent search...*, cit.