

Introduzione

Patiris è un servizio web progettato e realizzato da ricercatori del Dipartimento di Scienze Aziendali dell'Università di Bologna in collaborazione con UIBM (Ufficio Italiano Brevetti e Marchi) del Ministero dello Sviluppo Economico e con il supporto tecnico di Epoca Ricerca. I dati brevettuali sono stati raccolti attraverso la piattaforma Orbit.com. Tutti i nomi delle università italiane e degli istituti di ricerca mappati sono stati disambiguati con il supporto tecnico di Questel.

I dati sono illustrati attraverso report dinamici che consentano di apprezzare:

- l'andamento complessivo nel tempo e i principali trend;
- posizionamento rispetto ad aree tecnologiche e geografiche;

La sezione ricerca consente di effettuare ricerche sui dati brevettuali e restituisce risultati ordinati e raggruppati per famiglia di appartenenza. Cliccando su ciascun documento è possibile visualizzare un riassunto dei principali dati bibliografici e aprire il dettaglio. La finestra di dettaglio mostra ulteriori informazioni relative al documento selezionato e consente di accedere direttamente al profilo dello stesso documento sul Espacenet.

I dati raccolti partono dai brevetti registrati a titolarità singola o congiunta di Università, CNR ed ENEA per poter contare su informazioni pubbliche certificate. Il rigore della scelta rispetto alla qualità dell'informazione raccolta sconta due limiti noti da chi si occupa di questi problemi: la minore affidabilità delle informazioni più recenti in considerazione dei tempi tecnici di completamento delle procedure di registrazione; la non considerazione di brevetti a titolarità di soggetti diversi, ma che riportano come inventore un universitario o un ricercatore del CNR o dell'ENEA. Il primo aspetto è un tratto ineludibile legato all'uso di dati brevettuali. Il secondo comporta la risoluzione di problemi tecnici molto specifici di combinazione delle informazioni relative agli inventori e alle loro caratterizzazioni professionali. Su questo specifico tema rimandiamo al lavoro dei colleghi impegnati nel progetto APE-INV (Lissoni, Sanditov and Tarasconi 2006; Lissoni 2013).

I temi di discussione collegati alla brevettazione nelle istituzioni pubbliche sono molti e oggetto da tempo dell'attenzione di studiosi di diverse discipline, operatori privati e di decisori pubblici. In questo progetto non affrontiamo alcuno di questi, ma ci concentriamo nell'intento di fornire a tutti la base di dati necessaria a sostenere in maniera informata ed aggiornata tali studi e discussioni.

Metodi

1.1 Fonte Dati

Basandoci sull'esperienza accumulata con la prima versione del servizio web Patiris, sui suggerimenti ricevuti dagli utenti che lo hanno utilizzato¹ tra il 2008 e il 2010 e considerando potenziali sinergie e complementarità rispetto ai servizi offerti dagli operatori commerciali, abbiamo valutato le principali alternative e scelto come fonte dati la piattaforma Questel-Orbit. Tali sinergie e complementarità derivano dal rapporto instaurato tra NETVAL e Questel-Orbit. NETVAL raccoglie 49 università italiane, il Consiglio Nazionale delle Ricerche (CNR) e l'Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA) supportandone la valorizzazione dei risultati della ricerca attraverso attività formative e di networking con istituzioni, mondo delle imprese e della finanza. Nel contesto italiano, NETVAL rappresenta il 51,7% di tutti gli atenei italiani e circa il 95,8% dei brevetti universitari attivi (al 20/10/2009). La scelta della piattaforma Orbit, già ampiamente utilizzata dagli associati al consorzio NETVAL, aumenta la portata del progetto Patiris consentendo la potenziale condivisione di best practices sia in termini di data analysis che di ottimizzazione dell'esperienza utente relativamente alla pulizia del dato brevettuale per i documenti relativi alle Università e ai principali Centri di Ricerca pubblici italiani.

Orbit consiste in un archivio dinamico di dati organizzati e indicizzati che raccoglie più di 60 milioni di documenti brevettuali da oltre 92 autorità di tutto il mondo. Attraverso il database Fampat Orbit raggruppa i documenti per famiglie 'invention-based'. Di conseguenza, un singolo record per una famiglia brevettuale combina insieme tutte le fasi della pubblicazione della famiglia. Questa funzionalità è particolarmente utile, ad esempio, in considerazione delle richieste ricevute da diversi rappresentanti degli uffici di trasferimento tecnologico riguardo alla possibilità di avere una piattaforma online che preveda un profilo dedicato per ciascuna 'invenzione' - e non per ciascun documento - come strumento di promozione rispetto a possibili utilizzi dell'invenzione brevettata e alle competenze dell'istituto. Tra le alternative possibili, Orbit presenta alcune funzionalità adatte ad effettuare ricerche per gruppi di assignee, operazioni di pulizia del dato con raggruppamenti e de-duplicazione. Funzionalità che si rivelano utili per analisi longitudinali sul portafoglio brevettuale di gruppi e singoli EPR.

¹ Gli utenti potenzialmente interessati ad una mappatura del patrimonio brevettuale degli istituti di ricerca pubblica italiani sono: *professionisti* (consulenti di proprietà industriale, broker di tecnologie, ...); *Pubblica Amministrazione* (uffici di trasferimento tecnologico, camere di commercio, ...); *media e altro* (media, imprese, singoli individui). I professionisti utilizzano già un ampio ventaglio di servizi commerciali a supporto della propria operatività, hanno una conoscenza approfondita degli strumenti di proprietà industriale ed esprimono esigenze specifiche rispetto alle funzionalità e ai contenuti desiderati. Gli uffici di trasferimento tecnologico hanno mostrato interesse riguardo alla possibilità di avere un blog dedicato per ciascun documento come strumento di promozione rispetto a possibili utilizzi dell'innovazione brevettata. I media, infine, si sono mostrati molto interessati a rapporti periodici, analisi ed indicatori sulla produttività dei singoli enti pubblici di ricerca.

1.2 Unità d'analisi: Invenzioni

L'utilizzo di dati brevettuali per misurare l'attività innovativa necessita di precisi accorgimenti al fine di caratterizzare correttamente invenzioni piuttosto che documenti. In assenza di tali accorgimenti, la presenza di documenti brevettuali equivalenti può generare significative distorsioni.

In questo progetto non analizzeremo informazioni rilevanti ai fini di valutazioni di tipo legale, ma informazioni di natura principalmente bibliografica nell'intento di caratterizzare l'attività inventiva degli EPR italiani. Per questo motivo, piuttosto che focalizzarci su singoli documenti brevettuali, osserviamo gruppi di domande di brevetto depositate e/o brevetti concessi (in vari paesi e in diversi istanti temporali) relativi ad una comune invenzione. Tali gruppi di documenti sono definiti 'famiglie di brevetti'.

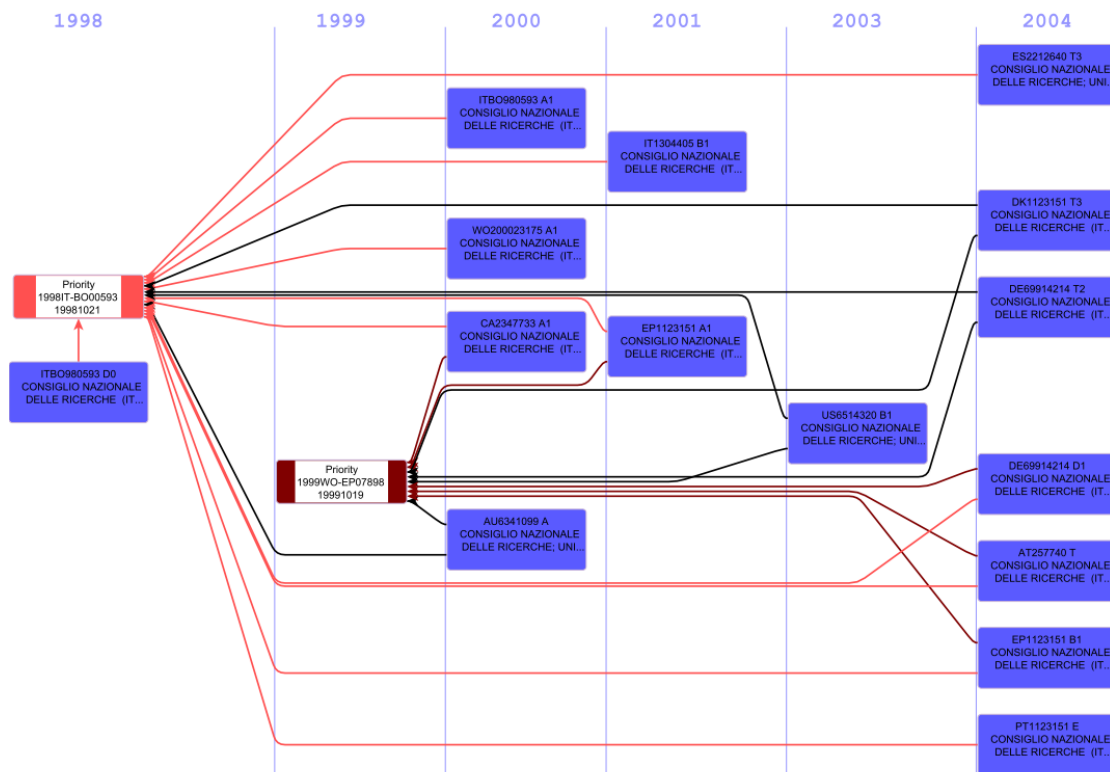
Con il termine famiglia di brevetti si intende generalmente un gruppo di documenti brevettuali che, come una famiglia, sono collegati fra loro. Il collegamento è rappresentato da una comune priorità. La data di priorità e il relativo numero di priorità indicati su tutti i documenti brevettuali indicano la data a partire dalla quale viene rivendicata la novità di un'invenzione e decorrono i diritti sulle domande depositate successivamente. Generalmente questa corrisponde alla prima data di deposito nella nazione di appartenenza del richiedente. I documenti appartenenti ad una stessa famiglia sono detti "equivalenti" e hanno in comune uno o più numeri di priorità.

Esistono diversi tipi di famiglie di brevetti contraddistinte dal criterio utilizzato per raggruppare i singoli documenti brevettuali. Ciascun tipo di famiglia presenta pro e contro a seconda del tipo di analisi che s'intende effettuare. L'insieme di questi aspetti rende la ricerca e l'analisi d'informazioni brevettuali attività complesse e particolarmente influenzate dalle scelte soggettive dell'analista.

La famiglia INPADOC² è definita come comprendente tutti i documenti che hanno la stessa priorità o combinazione di priorità. Questo include tutti i documenti brevettuali derivanti da una domanda di brevetto presentata come un primo deposito presso un ufficio brevetti e dalla stessa domanda di brevetto presentata entro l'anno di priorità presso un qualsiasi ufficio brevetti di un altro paese. Questo tipo di famiglia è particolarmente utile per effettuare analisi di prior art e stato legale poichè consente di trovare tutti i membri di una famiglia che sono anche solo lontanamente collegati, ma può risultare troppo ampia per il nostro obiettivo (individuare singole invenzioni), poichè può rendere difficile trovare l'esatta relazione di un singolo documento brevettuale con un altro.

² INPADOC (International Patent Documentation Center) è un database di documenti brevettuali pubblicamente accessibile. È gestito e aggiornato settimanalmente da EPO fornendo informazioni su "famiglie di brevetti" e sullo stato giuridico. Domande di brevetto in diversi paesi che citano la stessa priorità sono considerate come un'unica galassia di documenti a protezione di una comune invenzione. INPADOC fornisce inoltre informazioni relative allo status giuridico delle domande di brevetto e brevetti concessi in quei paesi che segnalano cambiamenti di stato.

Fig. 1 – Struttura di una Famiglia di Brevetti INPADOC



Sul fronte opposto la famiglia EPO in senso stretto prevede che due documenti possano essere considerati equivalenti se condividono tutti la stessa priorità. Si tratta di un tipo di raggruppamento per invenzione molto preciso ed accurato per strategie di analisi orientate all'individuazione di prior art, ma probabilmente troppo restrittivo nel caso di strategie di ricerca nominative, come quelle condotte per la costruzione del database alla base del presente progetto.

Le famiglie di brevetti "Derwent" consultabili attraverso il database Derwent World Patent Index (DWPI) di Thomson Reuters sono generate attraverso un processo di expert coding. Thomson Reuters utilizza esperti con competenze sviluppate in oltre 40 anni per creare famiglie di brevetti i cui membri hanno contenuti tecnici simili.

La definizione di famiglia utilizzata in questo progetto è la famiglia "FamPat", sviluppata da Questel. La famiglia FamPat espande la regola rigida della famiglia EPO con norme aggiuntive in modo da incorporare variazioni nella definizione di invenzione di diverse autorità brevettuali (es. particolarmente utile per la ricerca pubblicazioni giapponesi e statunitensi). La famiglia e il database FamPat combinano insieme tutte le fasi di pubblicazione dei documenti membri. Ricerche per assegnatario, inventore o classi sono perfezionate su tutti gli equivalenti della famiglia. Gli ID assegnati a ciascuna famiglia sono stabili nel tempo.

L'insieme dei brevetti che compongono una famiglia rappresenta documenti diversi che fanno capo allo stesso numero di priorità. Questo significa che nonostante una stessa invenzione sia sottoposta a discipline giuridiche differenti a seconda dei paesi in cui la domanda di brevetto viene depositata, dal punto di vista della valutazione della performance innovativa tutti i documenti di una famiglia identificano un comune prodotto della ricerca. Di conseguenza, il semplice conteggio dei brevetti concessi o

delle domande di brevetto, senza raggruppamento in famiglie dei singoli documenti, restituisce un indicatore parziale e non del tutto affidabile relativamente alla performance innovativa o specializzazione tecnologica di individui, imprese e nazioni.

Nel caso delle procedure EPO e PCT, ad esempio, una singola domanda di brevetto permette di ottenere a cascata una pluralità di documenti brevettuali al fine di proteggere l'invenzione sottostante in molteplici stati senza dover ripetere, stato per stato, tutte le singole procedure di domanda.

1.3 Anagrafica Istituti, Individuazione Varianti di Nomi

Il nostro obiettivo richiede di identificare tutte le famiglie di brevetti che hanno per assegnatario un'università o istituto di ricerca pubblico italiano con l'obiettivo di costruire rapporti periodici sulla brevettazione accademica. A questo scopo abbiamo definito un'anagrafica degli istituti monitorati basandoci sulle informazioni riportate dal Ministero dell'Istruzione, dell'Università e della Ricerca sul proprio sito istituzionale www.miur.it. Il database definitivo contiene 94 Università a cui si aggiungono Consiglio Nazionale delle Ricerche (CNR) e Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA) per un totale di 96 istituti.

Partendo dai nomi degli istituti abbiamo effettuato per ciascuno numerosi cicli incrementali di ricerche e raffinamento delle stringhe testuali con l'obiettivo di determinare gruppi di stringhe di ricerca ottimali ed individuando le possibili variazioni di nome osservabili nei documenti brevettuali.

Questa attività, fondamentale per consentire nel tempo ricerche stabili e coerenti in grado di restituire sempre gli stessi risultati, si è resa necessaria poiché non esiste un codice univoco assegnato agli istituti dalle diverse autorità internazionali. Al contrario, di volta in volta, gli istituti che figurano come assegnatari vengono identificati (es. al momento dell'applicazione) soltanto attraverso il nome e altre informazioni (es. indirizzo) codificate sotto forma di stringhe testuali. Di conseguenza, nell'effettuare ricerche basate sul nome dell'assegnatario vanno necessariamente risolti evidenti problemi di ambiguità relativi ai nomi di un istituto che vengono indicati nel corso degli anni sui documenti brevettuali.

A causa, ad esempio, di errori di data entry, variazioni di uno stesso nome o artefatti generati durante il processo di digitalizzazione di documenti cartacei tramite tecnologia OCR, si verificano numerosi casi in cui ad uno stesso istituto corrispondono molteplici varianti dello stesso nome. Per risolvere il problema è necessario costruire per tentativi stringhe di ricerca via via sempre più complesse al fine di catturare la varietà di nomi presenti nello stock di documenti brevettuali che s'intende osservare. Senza effettuare ricerche con operatori booleani per concatenare le possibili varianti risulterebbe impossibile identificare correttamente tutti i documenti e/o famiglie di documenti potenzialmente rilevanti.

1.4 Generazione Database Nomi Disambiguati, Definizione Stringhe di Ricerca

Controllando una ad una le varianti di nomi degli istituti individuate con il primo round di test e stringhe di ricerca concatenate abbiamo popolato un database che consente il raccordo tra i nomi 'standard' di università e centri di ricerca Italiani che risultano assignee di granted patents o patent applications e le varianti di nome individuate.

Il database finale contiene tutte le varianti presenti sui documenti brevettuali con accanto a ciascuna riportato il nome standardizzato MIUR. In media, per ciascuno degli istituti titolari di brevetti esistono 11 varianti sedimentatesi nel tempo per le ragioni più varie che rendono difficile sia l'analisi che la ricerca (es. 100 varianti per il CNR; 24 per l'Università di Bologna). Per illustrare il problema e la necessità di utilizzare un database per la disambiguazione dei nomi riportiamo qui di seguito due esempi reali con nome originale e nome standardizzato in parentesi:

UNIBERUSHITA DETSURI SUTOUDEI DAY CAGLIARI (Università degli Studi di Cagliari); EHNEHA EHNTÉ PER LE NUOVE TEKNOLODZHI L EHNERGIJA E L AMBIENTE (ENEA, Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile)

Partendo dai nomi degli istituti disambiguati abbiamo generato per ciascuno stringhe testuali per concatenamento con l'obiettivo di individuare la miglior stringa di ricerca sulla base delle variazioni di nome osservabili nei documenti brevettuali. Stringhe di ricerca ottimali, che risolvono completamente i problemi di name disambiguation, sono state individuate per tutti gli istituti italiani tranne gli atenei di Milano (per cui si è proceduto a controllare i documenti uno ad uno per assegnare i rispettivi nomi standardizzati).

In considerazione del beneficio potenziale per gli utenti Orbit associati al consorzio NETVAL derivante dalla disponibilità di nomi disambiguati, abbiamo raggiunto un accordo con Questel per l'ottimizzazione dell'esperienza utente relativamente alla pulizia del dato brevettuale per i documenti delle Università e dei principali Centri di Ricerca pubblici italiani. Per rendere fruibile al più ampio numero di soggetti interessati il lavoro di disambiguazione dai noi effettuato nell'ambito del progetto Patiris, i raggruppamenti e i nomi standardizzati da noi proposti sono stati sottoposti al Questel e dopo alcune settimane di test sono stati infine implementati sulla piattaforma online. Il nuovo raggruppamento da noi introdotto semplifica la ricerca dei brevetti degli istituti italiani rendendo possibile, per la prima volta, a chiunque utilizzi la medesima base dati, la semplice ricerca per nome standard MIUR (es. cercando "Politecnico di Milano" il sistema restituisce tutte le occorrenze associate a tutte le variazioni di nome che abbiamo verificato).

Credits

Sponsor

Loredana Guglielmetti

Ministero dello Sviluppo economico, Dipartimento per l'Impresa e l'Internalizzazione - Direzione Generale Lotta alla Contraffazione Ufficio Italiano Brevetti e Marchi – www.uibm.gov.it

Design & Development

Enrico Forti - <https://iris.ucl.ac.uk/iris/browse/profile?upi=EFORT49>

University College London, Department of Management Science & Innovation;
Università di Bologna, Dipartimento di Scienze Aziendali

Maurizio Sobrero -

<http://www.unibo.it/Faculty/default.htm?TabControl1=TabCV&mat=031026>

Università di Bologna, Dipartimento di Scienze Aziendali

Supporto Tecnico & Web Development

Fulvio Di Marco, Damiano Fontana, Mattia Generali

Epoca Ricerca - www.epocaricerca.it

Dati Brevettuali

Questel Orbit – www.orbit.com

Ministero dello Sviluppo economico, Dipartimento per l'Impresa e l'Internalizzazione - Direzione Generale Lotta alla Contraffazione Ufficio Italiano Brevetti e Marchi – www.uibm.gov.it

Supporto Tecnico (Integrazione Nomi Disambiguati in Database Questel)

Martine Massiera, Rossella Osella

Questel – www.questel.com

References

Lissoni F., B.Sanditov and G. Tarasconi (2006) "The Keins Database on Academic Inventors: Methodology and Contents", CESPRI working paper 181, Università "L.Bocconi", Milano.

http://www.francescolissoni.com/rp_g000004.pdf

Lissoni F. (2013) "Academic Patenting In Europe: A Reassessment of Evidence and Research Practices", *Industry & Innovation* 20/5: 379-384

<http://www.tandfonline.com/eprint/Mw42iPriUwu5CMS4D4q9/full>