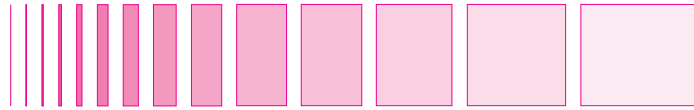




Camera di Commercio
Padova

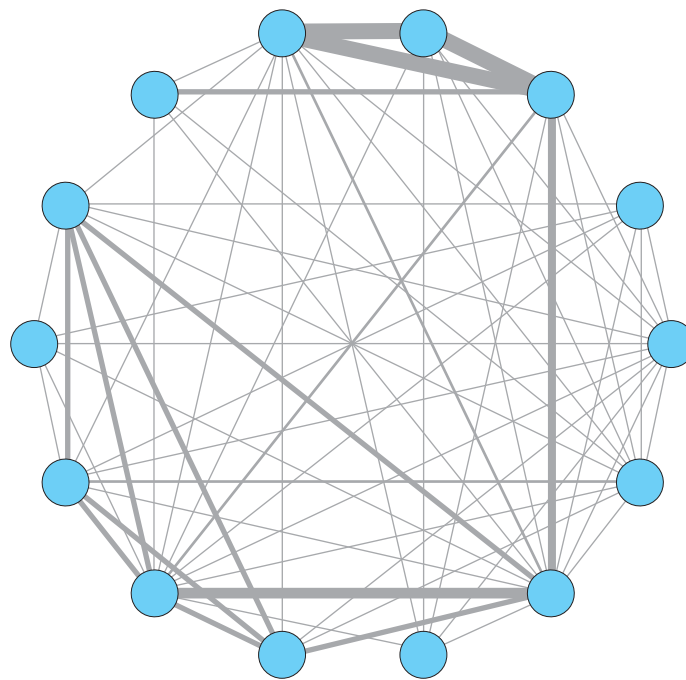
INNO-MED
innovazione nel medicale



**QUADERNI
di innovazione**

UN LESSICO DELLE COMPETENZE NEL MEDICALE

Executive Summary



Dall'indagine Preliminare
di Portafoglio brevettuale
e Posizionamento tecnologico
alla comprensione delle competenze
per una futura aggregazione





QUADERNI
di innovazione

INNOAED
innovazione nel medicale

UN LESSICO DELLE COMPETENZE NEL MEDICALE

Executive Summary



**Dall'indagine Preliminare
di Portafoglio brevettuale
e Posizionamento tecnologico
alla comprensione delle competenze
per una futura aggregazione**



a cura di
Paolo Nadalutti, Bep Srl
Bernardo Moronato, Bep Srl
Sandro Storelli, Area innovazione e ricerca, CNA Padova



La pubblicazione *“Un lessico delle competenze nel medicale”* è stata sviluppata nell'ambito di “INNO MED, Innovazione nel medicale”, progetto in compartecipazione tra Camera di Commercio di Padova e CNA di Padova.

Hanno curato la pubblicazione:

Paolo Nadalutti, Bep Srl

Bernardo Moronato, Bep Srl

Sandro Storelli, Area innovazione e ricerca, CNA Padova

Lavoro realizzato grazie alle ricerche del gruppo GIAT- Gruppo Interdisciplinare di Analisi Testuale, Università di Padova (<http://www.giat.org/>)

Comitato Tecnico Scientifico del progetto “INNO MED, Innovazione nel medicale”:

ANDRICH Renzo, Polo Tecnologico Don Gnocchi Milano - BERTO Vassilli,

Vassilli Srl - FRANCHIN Marco, Parco Scientifico Tecnologico Galileo -

GOMIERO Loretta, Officina Ortopedica Gomiero Srl - MARCATO Patrizio,

Ma.Vi.Dental Snc - MORONATO Bernardo, BEP Srl Consulting - PETRONE

Nicola, Dip. Ingegneria Meccanica Università Padova - PIVATO Gianfranco,

Rehateam Srl - POLO Federico, Osservatorio Biomedicale Veneto - RUSSO

Salvatore, Dip. di Management Università Ca' Foscari Venezia - SIMIONATO

Francesco, AIMAD Accademia Italiana Materiali dentali - STORELLI Sandro,

Osservatorio Biomedicale Veneto - TERRIN Vincenzo, L.O.R.I. Srl - TOSELLO

Domenico, Osservatorio Biomedicale Veneto - VAIA Giovanni, Dip. di Management

Università Ca' Foscari Venezia - VARIOLO Luigi, Ortopedia Variolo Snc -

VOLPATO Danilo, Off Carr Srl

Coordinamento del progetto:

TECNA Soc. Cons. a r.l.

Area Innovazione e Ricerca

Via della Croce Rossa, 56 - 35129 Padova

tel.049.8062236 fax 049.8062200

Grafica e copertina *Gianni Plebani*

Stampa a cura di *Il Prato srl*

© Tutti i diritti riservati:

CNA di Padova

CCIAA di Padova

Dicembre 2011

*Il linguaggio è la madre, non l'ancella
del pensiero.*
Karl Kraus

Presentazione

Il progetto *INNO MED - Innovazione nel medicale*, realizzato in compartecipazione tra Camera di Commercio e CNA Provinciale di Padova, ha articolato in logica di integrazione diverse linee progettuali di particolare interesse del comparto medicale.

Un importante obiettivo generale dato era quello di definire possibili formule collaborative per metodi condivisi tra Aziende pubbliche e imprese operanti nel settore di produzione e fornitura di dispositivi medici, nel comune intento di garantire percorsi di razionalizzazione e l'appropriatezza in sanità.

La definizione di modelli organizzativi condivisi e di procedure coerenti ed omogenee è infatti un obiettivo importante del sistema nel suo complesso.

Altro obiettivo era quello di analizzare le caratteristiche delle tecnologie applicate e/o di potenziale applicazione per lo sviluppo di prodotti innovativi con riferimento ai livelli di qualità prestazionale richiesti dal mercato, anche sviluppando coerentemente modelli e metodologie di determinazione dei costi del prodotto.

A questo fine è indispensabile considerare le specificità del prodotto - dispositivo medico, valutando oltre ai costi di produzione e distribuzione, i costi di servizio.

Strettamente collegato ai precedenti, era l'obiettivo di valorizzare le caratteristiche delle tecnologie e dei prodotti nelle loro caratterizzazioni prestazionali, nelle specializzazioni su base locale-territoriale, sviluppando anche un'analisi del trend di crescita degli asset intangibili nelle imprese.

A questo fine, è risultato importante il monitoraggio del trend tecnologico per lo sviluppo e la competitività di sistema.

Coerentemente a ciò, il progetto ha condotto un ulteriore approfondimento su filiere produttive locali individuate in sottosistemi di maggior interesse (maggiori prospettive di crescita, tecnologie più promettenti, coerenze con le prospettive di sviluppo regionale o altro, sviluppo della proprietà intellettuale) integrando risultati di studi sul sistema locale, sul rapporto tra economia pubblica e privata e tra tecnologia del territorio ed attività di ricerca finalizzata.

Altro obiettivo era quello di sviluppare tra le imprese consapevolezza sulle potenzialità di sviluppo di nuovi mercati, anche attraverso la proprietà intellettuale ed il trasferimento dell'innovazione.

Trasferire conoscenza alle imprese sul come sviluppare e proteggere nuovi prodotti in modo mirato sui mercati già presidiati ma in via di evoluzione e sviluppo, nonché su nicchie non presidiate o scarsamente presidiate, può aiutare ad anticipare la concorrenza.

Gli asset intangibili sono sempre più cruciali per lo sviluppo e la crescita.

Essi sono definiti come risorse che non hanno consistenza fisica ma il cui sfruttamento industriale ed economico produce benefici futuri.

In ogni caso, essi sono fattori di rilevanza crescente per la competitività.

Il vantaggio competitivo che risiede nel marchio, nelle innovazioni e nella protezione brevettuale, può diventare determinante anche per la PMI.

Valorizzare tali risorse è importante sia per capire meglio qual è il valore di ciò che possediamo, sia nell'ottica di trattative per cedere o dare in licenza i nostri asset, così come sviluppare attività in partenariato o inserire il valore aggiunto delle nostre risorse immateriali.

Di norma, la numerosità dei brevetti è ritenuto un importante indicatore della propensione all'innovazione.

La proprietà intellettuale può diventare un importante volano dello sviluppo economico di un territorio.

Quando il brevetto ha un successo effettivo sul mercato si realizzano anche apprezzabili ricadute nell'area interessata dallo sviluppo del prodotto.

Ciò avviene in termini di indotto, di trasferimento dell'innovazione, di crescita della cultura d'impresa a livello di filiera, di produzione della ricchezza.

Sono oggi in atto profonde mutazioni sul mercato:

- le nuove tecnologie applicabili producono modifiche radicali nei processi e prodotti, con una rapidità progressiva;
- nella società (e quindi nel mercato) crescono le richieste di prodotti di largo consumo, finalizzati al mantenimento dell'autonomia ambientale e soggettiva.

In questo contesto, è prevedibile un nuovo livello nella competizione, che rischia di vedere impreparate le imprese del territorio e di mettere a rischio per loro anche le quote di mercato interno.

Si tratta quindi di supportare le imprese nella crescita indispensabile di una cultura d'impresa capace di rispondere alle esigenze di un mercato aperto e potenzialmente in via di nuovo sviluppo.

Un approccio sistemico all'innovazione può offrire il supporto di più tecniche e metodologie, in grado di orientare il lavoro e riducendo la soglia di rischio.

Il supporto di servizio alle imprese nel territorio per la brevettazione è quindi di importanza strategica.

Va superata una diffusa "sottocultura" della proprietà intellettuale, più tendente a "datare l'ingresso sul mercato" del prodotto, più che ad attribuire valore all'idea progettuale.

A tal fine, diviene significativo mappare la proprietà intellettuale e la sua evoluzione, verificare l'adeguatezza dello specifico brevetto nel tempo, indirizzare per la finalizzazione sul mercato delle idee progettuali, analizzare le sinergie di ricerca applicata e progettazione tra diverse imprese e filiere, valutare l'incidenza della proprietà intellettuale sullo sviluppo del territorio regionale, comparare questa con altre realtà regionali.

Una prima indagine di merito è stata condotta sul comparto medicale, che si caratterizza nella realtà provinciale e regionale con un rapporto sinergico tra settori maturi e avanzati, ad alta tecnologia. Il lavoro di screening eseguito sulle aziende del biomedicale veneto, ha confermato come le dimensioni dei player siano mediamente molto piccole. Dall'analisi della proprietà intellettuale è emerso chiaramente come l'attività di ricerca e sviluppo delle grandi imprese sia in fermento, mentre i piccoli sono sempre all'inseguimento in aree di mercato che, spesso, potrebbero presidiare con maggiori profitti con il supporto di competenze complementari.

Diviene quindi opportuno dare logica prosecuzione alle attività progettuali svolgendo gli approfondimenti del caso, a partire dal campione imprese dell'area della provincia di Padova.

Nell'ambito di un progetto di ricerca condotto con un'Università leader in Europa, una società padovana ha messo a punto e testato una specifica metodologia per la valutazione economica di brevetti che rispetta le linee guida emanate nel 2008 da ABI, CRUI e Confindustria.

Questa metodologia si fonda sulla valutazione ponderata di una serie di parametri (ad es. il livello di "affollamento" tecnologico di un certo settore, oppure le potenzialità di mercato del prodotto/tecnologia legata al brevetto in esame).

Per rendere quanto più oggettiva la valutazione, ciascun parametro è determinato da dati ed informazioni estratti da banche dati professionali e successivamente elaborati. La citata metodologia è applicabile sia a singoli brevetti che a portafogli di brevetti e si presta ad un approccio “incrementale” alla valutazione di brevetti, praticabile anche per la piccola impresa.

Mappare un gruppo di aziende significa determinare per ciascuna di esse un insieme di parole chiave che definiscono in modo univoco:

- prodotti/servizi; mercati geografici e target di clientela;
- tecnologie e competenze; dimensione, fatturato o altri campi “finanziari”.

Mettendo in relazione tra loro aziende e parole chiavi è possibile valutare le aree di sovrapposizione e di complementarità: in pratica valutare il grado di integrazione attività propedeutica ad azioni di “Business Network”.

Le parole chiave sono estratte da un contenuto adatto rispetto all'obiettivo da raggiungere: contenuto di un sito web aziendale, presentazioni/documenti aziendali, Form appositamente predisposti.

Le parole chiave sono organizzate in categorie secondo un'ideale classificazione di riferimento o “tassonomia” (dal greco *taxis*, ordinamento e *nomos*, regola) allo scopo di avere coerenza logica e semantica. Le parole chiave possono essere estratte in modo automatico dal contenuto.

È opportuno mappare le tecnologie e competenze impiegate dall'azienda nel processo di generazione dei suoi prodotti e servizi; realizzare una classificazione partendo da quello che già esiste e andando a specializzare quanto indispensabile per un corretto inquadramento dell'azienda nella sua categoria

Gli obiettivi della mappatura sono:

- raggiungere un miglior grado di conoscenza reciproca al fine di poter mettere a fuoco più rapidamente e con maggiore consapevolezza azioni tipiche di “Business Network”;
- progettare alleanze di business basate sulla complementarità delle competenze, dei prodotti, dei servizi e dei mercati;
- individuare aree di investimento precompetitivo comune nelle quali impostare progetti di ricerca e sviluppo condivisi;
- miglioramento del rapporto Enti di ricerca-Impresa.

La mappatura delle aziende è finalizzata a costruire uno strumento di matching per:

- costruzione partenariato ideale per un bando d'incentivo all'innovazione o alla ricerca;
- identificazione certa delle aziende in diretta concorrenza (stesso prodotto - stesso target) o pienamente complementari (diverso prodotto - diverso target);
- manifestazioni di interesse di fornitori/clienti interessati a specifici prodotti.

Nel corso del 2011 e nell'ambito di Inno Med, abbiamo potuto svolgere una parte, crediamo significativa, di questo lavoro, analizzando su un campione d'impres e trend tecnologici il lessico più utilizzato, per definire percorsi e strumenti utili per possibili scambi e collaborazioni tra imprese, su questo territorio e sul mercato globale.

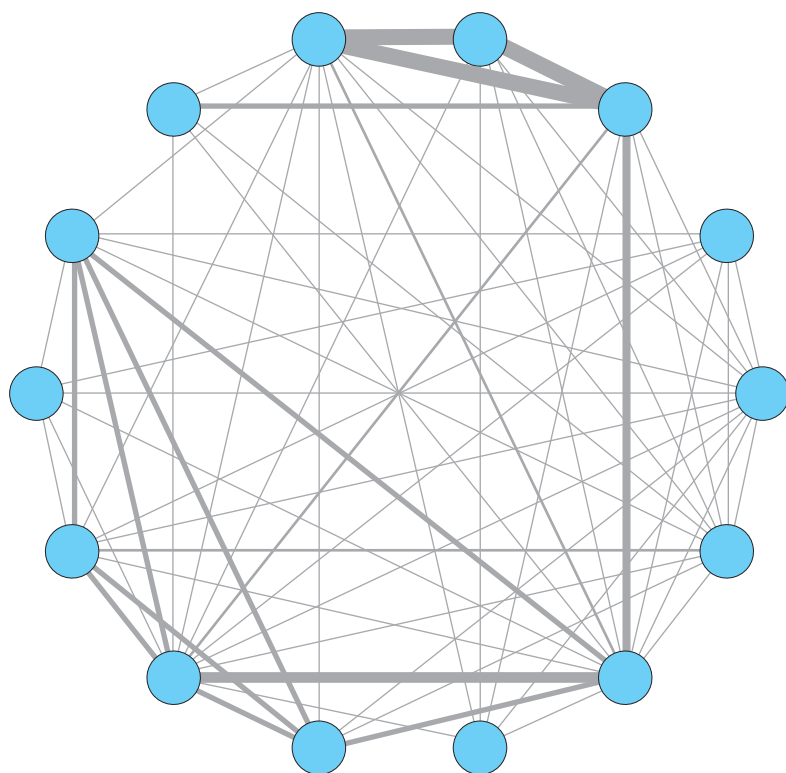
Contiamo che i risultati di questo lavoro, che pubblichiamo, possano essere utile presupposto per sviluppare ulteriormente le attività di mappatura nel medicale e non solo, per una nuova capacità competitiva per le imprese e per l'intero sistema regionale del Veneto.

Sommario

Introduzione	8
I TASSONOMIE E ONTOLOGIE	9
I.a Classificazione di documenti	9
I.b L'analisi quantitativa dei testi	10
II IL REPERIMENTO DEI DATI	11
II.a Modalità di ricerca di anteriorità	11
II.b Le basi dell'analisi testuale	11
III LA TECNOLOGIA NEL SETTORE BIOMEDICALE VENETO	13
III.a Il lessico più utilizzato	14
III.b Le aziende coinvolte	15
IV UNA TASSONOMIA PER IL SETTORE BIOMEDICALE	19
V ALCUNI TREND TECNOLOGICI	21
V.a Un altro esempio	22
VI CONCLUSIONI E POSSIBILI SVILUPPI	23



UN LESSICO DELLE COMPETENZE NEL MEDICALE



Introduzione

Lo scorso anno abbiamo realizzato un'indagine conoscitiva preliminare del portafoglio brevettuale e del posizionamento tecnologico delle aziende del Biomedicale Veneto. L'obiettivo che ci eravamo posti era la comprensione della vicinanza e del reale posizionamento tecnologico delle imprese del distretto. La principale domanda che ha guidato il nostro lavoro è stata: le aziende hanno requisiti sufficienti per poter "fare squadra"? Quante sono in competizione/concorrenza e quante potrebbero, secondo il posizionamento tecnologico emerso, lavorare assieme costituendo, di fatto, "massa critica" per avere un peso ed una penetrazione diversa nel mercato? Il risultato del lavoro ha fatto emergere, oltre alla polverizzazione delle specializzazioni tecnologiche, anche scenari di possibili collaborazioni, dovute alle competenze trasversali delle varie imprese. Proprio la polverizzazione delle competenze unita alla piccola-piccolissima dimensione delle aziende (due elementi che di solito costituiscono un limite per lo sviluppo), se assumiamo un diverso punto di vista, si trasformano da vincoli in opportunità: l'unione (la squadra) delle singole competenze e conoscenze, se attivate in modo sinergico, sono un patrimonio enorme che ad oggi non è ancora utilizzato.

Crediamo fortemente nella strada tracciata lo scorso anno, e il nostro cammino per il 2011 è la prosecuzione logica di quanto realizzato. Se fino a ora il punto focale della nostra attenzione si è concentrato sulle caratteristiche "esteriori" dei brevetti, per questa fase abbiamo voluto elaborare un sistema di analisi più approfondito, che portasse in luce non solo le diverse etichette dei brevetti, ma anche in grado di estrapolare i contenuti di questi preziosi testi.

Abbiamo creato un "sistema di misura" dei contenuti tecnici (tassonomia) tramite analisi di tipo semantico dei brevetti. Il nostro scopo finale è di portare in evidenza le competenze delle aziende e magari spingerci più in là, con l'ambizione di poter elaborare uno strumento informatico che permetta l'identificazione delle competenze delle aziende, così da poter individuare le possibili collaborazioni nei diversi aspetti della produzione. Non si tratta delle Pagine Gialle, questo è molto di più e potrà essere uno strumento trasversale per le imprese del Distretto Biomedicale per "fare squadra", per recuperare le risorse necessarie alla realizzazione di un prodotto, per risolvere un problema o per poter correre assieme ad una gara con maggiori possibilità di successo.

Questo nuovo strumento è al momento zero della sua esistenza e necessita di correzioni e collaborazione da parte delle aziende per poter crescere ed essere efficiente ed efficace per tutti. Secondo noi questo nuovo sistema ha tutti i presupposti per diventare il mezzo dell'inizio della collaborazione tra piccole imprese che vogliono crescere.

I Tassonomie e ontologie

La pubblicazione *Medicale veneto e brevetti, indagine preliminare di portafoglio brevettuale e posizionamento tecnologico* (di seguito *Medicale veneto e brevetti*), pubblicata nel 2010 dall'Osservatorio Biomedicale Veneto, tracciava una prima analisi delle tecnologie presenti in Regione per quel che riguarda il settore biomedicale.

L'analisi riguardava i dati presenti in 980 brevetti che orbitavano attorno alle aziende regionali, sia per quel che riguarda la produzione intellettuale interna (cioè brevetti depositati da aziende della regione), sia per quel che riguarda l'essere "satelliti" di aziende estere da parte delle aziende nostrane (cioè brevetti di aziende estere che hanno sedi anche in Veneto).

La capacità di produrre tecnologie è una dimensione del tessuto produttivo difficilmente rilevabile. Le tecnologie sono difficilmente "separabili" l'una dall'altra. Esistono istituti internazionali che hanno per scopo primario la separazione in "rami di conoscenza" delle tecnologie. È inoltre difficile andare a selezionare le aziende coinvolte in un determinato settore, vuoi per la mobilità delle imprese, vuoi per la non completezza delle informazioni che a volte dimostrano le classificazioni ufficiali. Un altro elemento di difficoltà sono le modalità di indagine: le metodologie classiche dovrebbero arrivare a risultati certi, ma sappiamo che il settore biomedicale è frammentato, e più sono frammentati i campi di indagine, meno sono precisi gli esiti; come mettere a punto un questionario che sia soddisfacente sia per chi si concentra sull'estetica che per chi si concentra sulla salute? Senza contare che interviste e questionari non vanno che a sommarsi alle già molte attività di rilevazione dati condotte in regione (ISTAT, Camere di Commercio, Regione, etc.). Per ultima questione (ma non meno importante) al momento non c'è modo di

fondere in un solo quadro descrittivo i capitali tecnologici delle aziende e quelli umani: ricordiamo che le imprese sono innanzitutto frutto delle persone che le hanno create, e a loro immagine veicolano anche valori, emozioni, esperienze.

Per riassumere, condurre un'indagine sul territorio può rivelarsi:

- Difficile
- Inadeguato
- Incerto nei risultati

In questo senso, la possibilità fatta intravedere dalla pubblicazione *Medicale veneto e brevetti*, è una novità assoluta per un bacino di utenze come quello Veneto. In questo studio ci proponiamo di dare continuità a questo approccio, andando a sondare più in profondità i contenuti del portafoglio brevettuale regionale.

■ I.a Classificazione di documenti

Abbiamo già accennato alla numerosità dei brevetti a nostra disposizione (980). In aiuto alle analisi che altrimenti dovremmo condurre manualmente (leggendo!) possono venire metodi automatici di classificazione.

Gli Uffici Brevetti organizzano le collezioni di documenti in tassonomie. La più importante di esse è l'IPC, l'International Patent Classification. La World Intellectual Property Organization (WIPO) cura l'aggiornamento continuo della IPC utilizzando un supporto XML che utilizza le tassonomie come concetti e le citazioni vicendevoli dei documenti come relazioni. La tassonomia dell'IPC consiste in circa 80.000 categorie che coprono l'intera gamma delle tecnologie industriali. Ci sono otto sezioni denominate dalla A alla H al più alto livello della gerarchia, dopodiché ci sono 128 classi, 648 sotto-classi, 7200

gruppi principali e 72.000 sotto-gruppi ai livelli inferiori.

A contribuire a queste classificazioni, negli ultimi anni sono intervenuti molti strumenti automatizzati che, conosciuti sotto vari nomi, confluiscono essenzialmente negli studi dell'AI (*artificial intelligence*). Parliamo delle tecniche di *document clustering*, del *natural language processing*, del *text mining*. La costruzione di basi di conoscenza molto vaste e che coprano numerosi argomenti è un tema cruciale nell'intelligenza artificiale, ma l'enorme difficoltà incontrata nella compilazione manuale di ontologie sufficientemente vaste e la rapidissima crescita del web e delle banche dati online in termini di quantità di informazioni contenute, ha portato molti a cercare delle tecniche per riversare le informazioni disponibili in ontologie utilizzabili dalle macchine.

Per il settore biomedicale veneto abbiamo intravisto la possibilità di creare una tassonomia non di brevetti, ma di aziende. I criteri di associazione delle aziende non saranno a questo punto le vicendevoli citazioni dei testi, ma la vicinanza in termini di argomenti trattati: con questa innovazione potremo mappare non solamente tecnologie, ma anche cultura di impresa e immagine.

■ I.b L'analisi quantitativa dei testi

La realizzazione delle tassonomie è stata fortemente accelerata dall'informatica. Ma l'informatica non si è mossa spinta solamente dalle necessità del settore industriale. Le prime applicazioni delle tecniche che sono poi state riprese dallo studio delle AI derivano in parte dalla linguistica quantitativa e dalla statistica. Alla materia hanno contribuito anche sociologi e psicologi, ad esempio utilizzandole per particolari analisi di interviste (se condotte su un alto numero di soggetti).

Il minimo comun denominatore di que-

ste tendenze è la trasformazione dei testi scritti in dati numerici, e perciò l'individuazione di indici di similarità tra testi, indici di utilizzo lessicale, misure di pertinenza dei testi con argomenti dati.

Il web moderno con i suoi potenti motori di ricerca non ha fatto altro che riprendere i concetti che abbiamo esposto in questi brevi paragrafi: ha coniugato l'esigenza di creare enormi classificazioni delle pagine multimediali con la possibilità di analisi dei comportamenti umani. Internet ha dato una grossa spinta alla materia, dandole l'occasione di giungere a una nuova maturazione coniugando le visioni "sociali" e quelle informatiche, convertendo quelli che sono stati per anni strumenti in mano a poche realtà sparse, a potenti strumenti di analisi della realtà.

Come abbiamo associato questi pensieri alle realtà aziendali? Proviamo solamente a pensare a quanti dati numerici affidabili sono disponibili attorno a noi, e quanti sono invece i testi. Secondo alcune stime l'80% delle informazioni che transitano nelle aziende sono in forma di testi, e non di numeri. Inoltre, confrontare i dati "scritti" relativi a molte aziende diventa difficoltoso, lungo e poco affidabile, perché dal momento in cui si comincia l'analisi al momento in cui si termina può trascorrere molto tempo, e le ipotesi di analisi iniziali possono anche trasformarsi. È anche per questo che sono nate le tecniche di analisi quantitativa del linguaggio.



II Reperimenti dei dati

L'obiettivo della ricerca brevettuale è l'estrazione e l'analisi strutturata di informazioni relative a soluzioni nel campo applicativo di interesse. Il reperimento dei vari documenti brevettuali necessita inevitabilmente di una serie di step sequenziali, all'interno dei quali saranno individuate delle procedure standardizzate da seguire. Ciò permette un affinamento degli obiettivi da raggiungere e favorisce una migliore qualità dell'analisi.

■ II.a Modalità di ricerca di anteriorità

La pubblicazione *Medicale veneto e brevetti* portava alla luce per il Veneto un portafoglio brevettuale di 980 testi, andando ad individuarli tramite le procedure standard utilizzate dai documentaristi brevettuali. Per le finalità del lavoro è risultato opportuno eseguire una successione ben precisa di passaggi.

Si è verificato, per prima cosa, che il sistema in esame avesse delle classi brevettuali di riferimento, affinando la scelta per passi successivi e andando a valutare la pertinenza delle stesse. La selezione di tali classi è passata attraverso l'immissione di termini inglesi (e relativi sinonimi) corrispondenti all'oggetto della ricerca nel relativo campo della *classification search*. Identificando una o più specifiche classi della tecnica da monitorare, si limita la ricerca al loro interno. Una volta identificata la classe o le classi di pertinenza, queste verranno usate come filtro per le ricerche brevettuali future inserendole, se necessario e non troppo vincolante, nella stringa relativa alla classe brevettuale IPC.

Non sono stati fornite indicazioni per quel che riguarda l'intervallo di tempo all'interno del quale ricercare, dando per obiettivo il monitoraggio del capitale tec-

nologico accumulato.

Per ultimo, sono state scelte le parole-chiave riguardanti le tecnologie (da ricercare nel titolo, abstract, rivendicazioni od in entrambi), le aziende titolari dei brevetti, l'inventore/i e, ove necessario, scegliere la fonte di estrazione dei brevetti.

Per il presente lavoro la ricerca brevettuale è stata semplicemente aggiornata impostando direttamente nei criteri di ricerca i nomi delle aziende emerse e i codici dell'IPC relativi al settore.

Non tutti i testi si prestano ad essere analizzati: dei 980 brevetti a disposizione non si dispone del testo completo di una parte di essi (disponibili solamente su richiesta agli Uffici Brevettuali), alcuni altri sono scritti in lingue straniere. Si analizzano quindi i testi di 516 brevetti, relativi a 108 aziende.

■ II.b Le basi dell'analisi testuale

I testi dei brevetti non possono essere analizzati nella forma grezza in cui si presentano. Dato che molte analisi si basano sul conteggio delle parole che compaiono nei diversi documenti (approccio *bag of words*), dobbiamo trattare la collezione di brevetti (il corpus) per renderlo omogeneo e analizzabile. Le analisi sono state condotte grazie al pacchetto tm del software statistico R¹. Gran parte delle routine sono comunque frutto di programmazione autonoma condotta all'interno della BEP e del gruppo di ricerca GIAT² dell'Università di Padova.

¹ R Project (cran.r-project.org)

² GIAT - Gruppo Interdisciplinare di Analisi Testuale (<http://www.giat.org>)

L'informazione testuale viaggia su binari differenti a seconda delle tipologie di testi. Nel caso dei brevetti esiste una già solida esperienza di trattamento, che ha codificato i passaggi di selezione delle unità di analisi di base. Secondo questa metodologia i termini che costituiscono la base informativa sono pochi e molto descrittivi. Le unità su cui ci baseremo sono essenzialmente sostantivi e verbi, che vanno presi nel loro significato più generale; vanno considerati come una sola parola i singolari e i plurali di uno stesso nome, e tutte le diverse coniugazioni dei verbi. Ad esempio, se parlassimo di romanzi, dovremmo tenere sicuramente in conto congiunzioni e altri termini "vuoti di significato", che vanno a caratterizzare uno stile di scrittura.

Meno tipica di questo approccio è l'individuazione dei segmenti di testo. I segmenti sono porzioni di testo, relativamente brevi - da 2 a 10 parole -, che si ripetono con continuità nel corpus. Tipici esempi sono "acido ialuronico", oppure "sedia a rotelle".

L'ultimo passo è eliminare tutte quelle parole che sono "troppo rare" per risultare utili alla classificazione di un brevetto: per le analisi si considerano solamente le parole che si ripetono un determinato numero di volte.

Come è facile intuire il database brevettuale subisce parecchie trasformazioni, che lo rendono quasi "illeggibile". Nella fattispecie, i passi sono i seguenti:

1. caricamento dei dati su R;
2. eliminazione delle parole relative al lessico convenzionale brevettuale ("claim", "claims", "apparatus", "device", etc.);
3. abbassamento delle maiuscole e eliminazione di punteggiatura e numeri;
4. *stemming* del testo: procedimento che elimina tutti i suffissi delle parole. Più precisamente vengono mantenute solamente le parti iniziali, in modo da rendere uguali singolari e plurali e diverse coniugazioni dei verbi;

5. individuazione dei segmenti di testo ripetuti con frequenza superiore o uguale a 10 e lunghezza massima 4;

6. Eliminazione delle *stopwords* (parole prive di significato compiuto in sé, come congiunzioni, preposizioni, etc.);

7. Eliminazione delle parole che non si ripetono almeno 40 volte nel corpus.

Alla fine di questo procedimento sono relativamente poche le parole che rimangono a descrivere un documento. Per delineare i profili dei testi (e dunque delle aziende che li hanno scritti) ci baseremo su circa 2.000 parole di significato "pieno" che presentano un numero adeguato di occorrenze, utilizzate nei brevetti delle 108 aziende.

Il vocabolario originale ha una dimensione di 938.849 parole di cui 25.637 diverse, per una ricchezza lessicale di 0,027. Le parole "uniche", i cosiddetti hapax, sono il 41% del vocabolario. Questi dati denotano un utilizzo del lessico molto peculiare.



III La tecnologia nel settore biomedicale veneto

Figura 1 e Figura 2 sintetizzano la situazione "pre analisi semantica" delle informazioni disponibili sul corpus di brevetti, e illustrano i collegamenti presenti tra le aziende basandosi sulle classificazioni IPC. Sono state ottenute andando a collegare tra loro le aziende che pubblicano brevetti nello stesso settore dell'IPC. Ci proponiamo di fare un passo avanti, e di verificare quali siano le competenze puntuali delle aziende di queste mappe. Esistono due diverse "zone tecnologiche": quella illustrata in figura 1 può essere definita "ad alta densità", in quanto composta da

aziende che brevettano in molti settori della IPC, e pertanto sembrano molto simili le une alle altre (lo spessore delle linee indica l'intensità della similarità). Mentre in figura 2 le aziende sembrano molto "sparse", ma in verità semplicemente pubblicano in un numero minore di settori dell'IPC. Notiamo subito come le varie "divisioni" della Fidia S.p.A. e della Orthofix siano correlate tra loro, ma quel che necessita di essere approfondito in queste due figure è la tipologia di settore in cui le aziende sono simili, ovvero le competenze che rendono diverse le aziende.

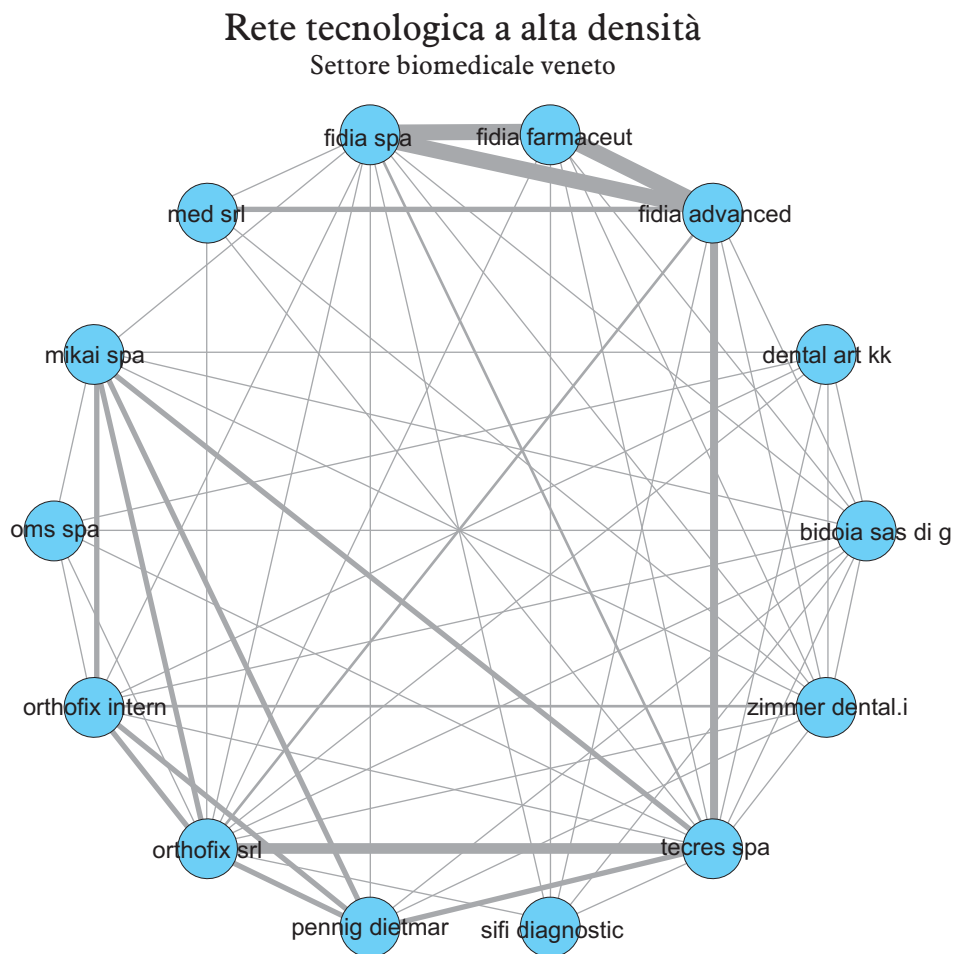


Figura 1: rete tecnologica basata sulla IPC

Rete tecnologica a bassa densità

Settore biomedicale veneto

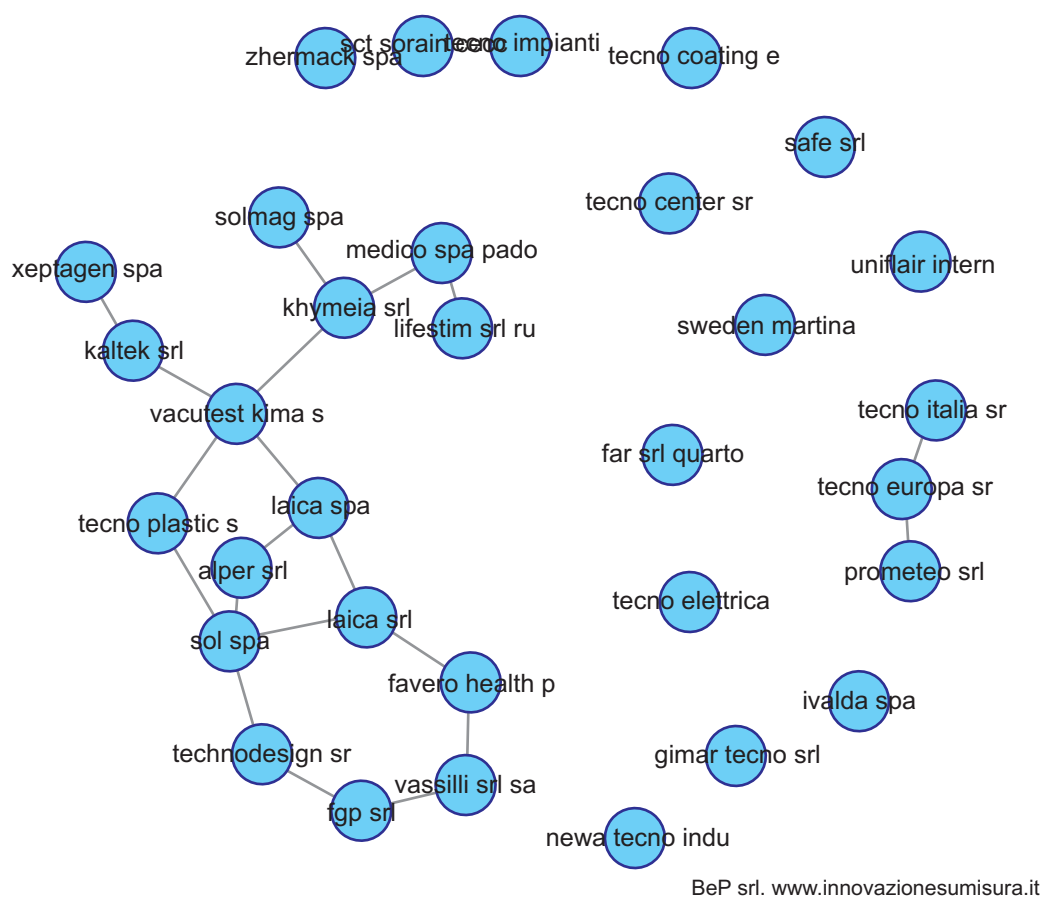


Figura 2: rete tecnologica basata sulla IPC

III.a Il lessico più utilizzato

Abbiamo già intuito come le procedure del *text mining* si basino essenzialmente sul conteggio delle parole. Comunque non tutte le parole, anche se molto utilizzate e perciò molto descrittive, sono utili allo stesso modo. Ci sono parole che sono molto utilizzate da alcuni gruppi di documenti, e molto poco utilizzate da altri, e sono queste quelle che effettivamente producono informazione. Noi ci aspettiamo che l'utilizzo delle parole cambi in base all'azienda che ha depositato il brevetto e, in misura minore, in base all'anno di deposito del brevetto.

Esiste una vasta gamma di tecniche statistiche per valutare l'effettivo sovra o sotto-utilizzo del lessico. In questa pubblica-

zione, per le analisi preliminari si utilizzerà il *calcolo delle specificità*, che si basa su modelli statistici parametrici di analisi³. Un'altra tecnica molto comune per l'individuazione del lessico più informativo è l'*indice TFIDF*: questo indice mette in evidenza le parole che non hanno una distribuzione uniforme nel corpus. Più una parola è concentrata in un numero ristretto di documenti, maggiore è l'indice TFIDF associato ad essa.

³ Si ipotizza che il numero delle occorrenze segua un modello multinomiale. La probabilità calcolata utilizza come parametri la dimensione del corpus, il numero di occorrenze per azienda e il numero di occorrenze della parola in esame.

La Tabella 1 fornisce un esempio del lessico più utilizzato nel corpus brevettuale a nostra disposizione. Buona parte dei termini individuati appartengono al mondo dell'industria farmaceutica di base, che, come abbiamo visto nella pubblicazione *Medicale veneto e brevetti*, è tra le più prolifiche in termini di produzione brevettuale. Tra i termini riportati possiamo notare la presenza anche dell'industria ortodontica. Evidentemente questo elenco risente del gran numero di brevetti di questi campi, ma il biomedicale non si esaurisce qui. Da un punto di vista statistico alcuni termini risultano particolarmente efficaci nel descrivere sottogruppi di brevetti. Ciò è dovuto alla loro diffusione limitata e molto concentrata in un numero limitato di aziende. I termini di Tabella 2 sono stati selezionati in modo da far emergere quelli che identificano con più precisione i settori tecnologici di cui fanno parte, utilizzando l'indice TFIDF. Un'azienda che attua una politica tecnologica avanzata è contraddistinta da elementi culturali e economici che si riflettono nell'utilizzo di un lessico del tutto peculiare e che non fa parte del patrimonio di conoscenze delle altre aziende. Un modo di individuare singolarità e sovrapposizioni tecnologiche è analizzare le aree lessicali comuni, ovvero individuare le aziende che condividono parti sensibili di vocabolario.

■ III.b Le aziende coinvolte

Veniamo al cuore della pubblicazione: il confronto tra i lessici utilizzati dalle aziende. La tabella 3 mostra un estratto del lessico più utilizzato dalle aziende che hanno il patrimonio brevettuale più sviluppato. La tabella è divisa in due parti, e confronta le parole che sono utilizzate più di frequente con quelle che lo sono in modo statisticamente significativo, secondo lo studio della significatività.

In tabella 3 i termini sottolineati in rosso, per quanto non siano quelli in valore asso-

luto più frequenti per l'azienda, sono più frequenti se confrontati con l'utilizzo che ne fanno le altre aziende. Andiamo così a scoprire ad esempio che le tre aziende che fanno riferimento alla Fidia Spa si specializzano in settori differenti: la Fidia Advanced Biopolymers in composti chimici di base, la Fidia Farmaceutici in farmaci antitumorali e applicazioni del collagene, la Fidia Spa fa esplicito riferimento a malattie, e pertanto alle applicazioni dei farmaci.

Prendendo spunto da questo approccio, possiamo anche associare, ad ogni termine, l'azienda che lo utilizza più spesso. Il metodo di lettura è immediato, e può restituire a determinate ricerche effettuate sul database lessicale, il nome dell'azienda che con maggiore probabilità è associata a questi termini.

Termine	Frequenza assoluta	Termine	Frequenza assoluta
acid	8.955	temperature	1.730
ml	7.804	solvent	1.677
solution	4.952	salt	1.664
hyaluronic	4.478	tissue	1.638
mg	4.068	acetone	1.634
implant	3.988	reaction	1.621
bone	3.477	axis	1.596
water	3.440	patient	1.457
ester	3.168	structure	1.440
figure	3.102	molecular	1.426
material	2.971	nail	1.371
mixture	2.870	carboxy	1.341
derivatives	2.612	alcohol	1.301
body	2.374	shape	1.278
preparation	2.243	minutes	1.248
sodium	2.210	abutment	1.214
cells	2.159	reference	1.213
esters	2.106	organic	1.206
hours	2.017	washed	1.185
screw	2.003	guide	1.103
weight	1.903	polymer	1.048
dental	1.894	pharmaceutical	992
following	1.824	human	952
time	1.781	head	912
dried	1.764	atoms	897

Tabella 1: le 50 parole più frequenti nel corpus brevettuale



Parola	Indice TFIDF	Frequenza assoluta	Numero di documenti
wheelchair	2.24	227	11
ray	1.93	245	41
bed	1.91	223	30
arm	1.86	350	52
decorating	1.86	704	17
seat	1.85	514	81
roller	1.83	380	22
tip	1.81	227	42
brace	1.81	359	19
belt	1.80	256	15
screen	1.79	264	30
user	1.79	386	68
filtering	1.79	285	34
cartridge	1.78	503	29
knee	1.77	510	35
upright	1.77	224	26
supporting	1.77	430	102
articulated	1.76	425	47
lid	1.75	201	26
vertical	1.75	356	82
printing	1.74	302	19
front	1.74	380	119
orthopedic	1.74	221	42
passage	1.72	287	79
arranged	1.72	507	120
valve	1.70	349	22
connecting	1.69	309	90
designed	1.68	370	117
fixing	1.67	280	70
electrodes	1.67	227	19
limb	1.67	230	46
shaft	1.67	299	46
root	1.65	237	44
bearing	1.65	212	60
ear	1.65	212	16
slot	1.65	228	53
operating	1.64	314	116
aim	1.64	224	102
ceramic	1.64	696	52
wire	1.64	208	36
cylinder	1.64	223	42
ball	1.63	344	35
positioned	1.62	536	141
cutting	1.62	270	40
slide	1.62	235	60
cavity	1.62	710	110
tubular	1.62	608	64
plug	1.61	230	27

Tabella 2:
parole più
frequenti nel
corpus, ordinate
per indice TFIDF,
con frequenza e
numero di
documenti che le
contengono

Termini specifici					
fidia advanced biopolymers srl	fidia farmaceutici spa	fidia spa	Vassilli Srl	Zhermack Spa	Orthofix srl
amid	amin	acid	aeroplane	archimedean	adapt
antisolv	antitumor	acyl	airbus	argument	anchor
arylaliphatic_cycloaliphatic	carboxyl	ag	airopplanes	arguments	articul
biological_materi	collagenas	alicycl	ameliorative	axisymmetric	ball
biomateri	cytostat	aliphatic	appears	bad	bar
biomed	deacetyl	alzheimer	appendance	band	bone
cell	drug	chronic	approched	bandwidth	drill
heavy_metal_salt	enzym	controlled_machin	arm	beliefs	elbow
hyaluronic_acid	hyaluronic_acid	epilepsi	armchair	bi	endoss
hyaluronic_acid_deriv	molecular_weight	etherifi	armchairs	bicomponent	guide_bor

Termini a frequenza di utilizzo più alta					
fidia advanced biopolymers srl	fidia farmaceutici spa	fidia spa	Vassilli Srl	Zhermack Spa	Orthofix srl
acid	acid	acid	seat	materi	screw
adhes	alcohol	acyl	wheelchair	impress	bone
agent	aliphatic	agent	frame	mix	clamp
alcohol	amid	alcohol	position	compon	hole
aliphatic	carboxyl	aliphatic	element	dispens	nail
amid	deacetyl	alkyl	wheel	object	portion
biomateri	drug	antibodi	vertical	exampl	distal
carboxi	enzym	carboxi	connection	mixer	fractur
cell	hyaluronic_acid	chronic	lever	colour	shape
hyaluronic_acid	molecular_weight	hyaluronic_acid	arm	devic	head

Tabella 3: parole statisticamente specifiche per azienda

IV Una tassonomia per il settore biomedicale

Finora abbiamo utilizzato le informazioni incorporate nei testi brevettuali per andare a evidenziare quali contenuti fossero più o meno peculiari di determinate categorie che conoscevamo a priori.

Andremo invece ora a "far parlare da se" i documenti: abbiamo visto come la tecnologia dipenda strettamente dal lessico. Il nostro scopo sarà ora raggruppare le aziende in base al lessico che utilizzano.

Per rendere l'idea del tipo di tassonomia che è possibile effettuare in base al lessico utilizzato possiamo andare a misurare la vicinanza di due aziende in base ai brevetti pubblicati: per dire se due aziende appartengono a uno stesso gruppo ricorremo a misure di "distanza lessicale" tra aziende. Quando la distanza lessicale tra due aziende è bassa, allora queste due aziende appartengono ad uno stesso gruppo. Questa implementazione dei raggruppamenti consente di conservare una caratteristica fondamentale di una tassonomia: la struttura gerarchica dei raggruppamenti. Ogni azienda entra a far parte di un "ramo" tecnologico.

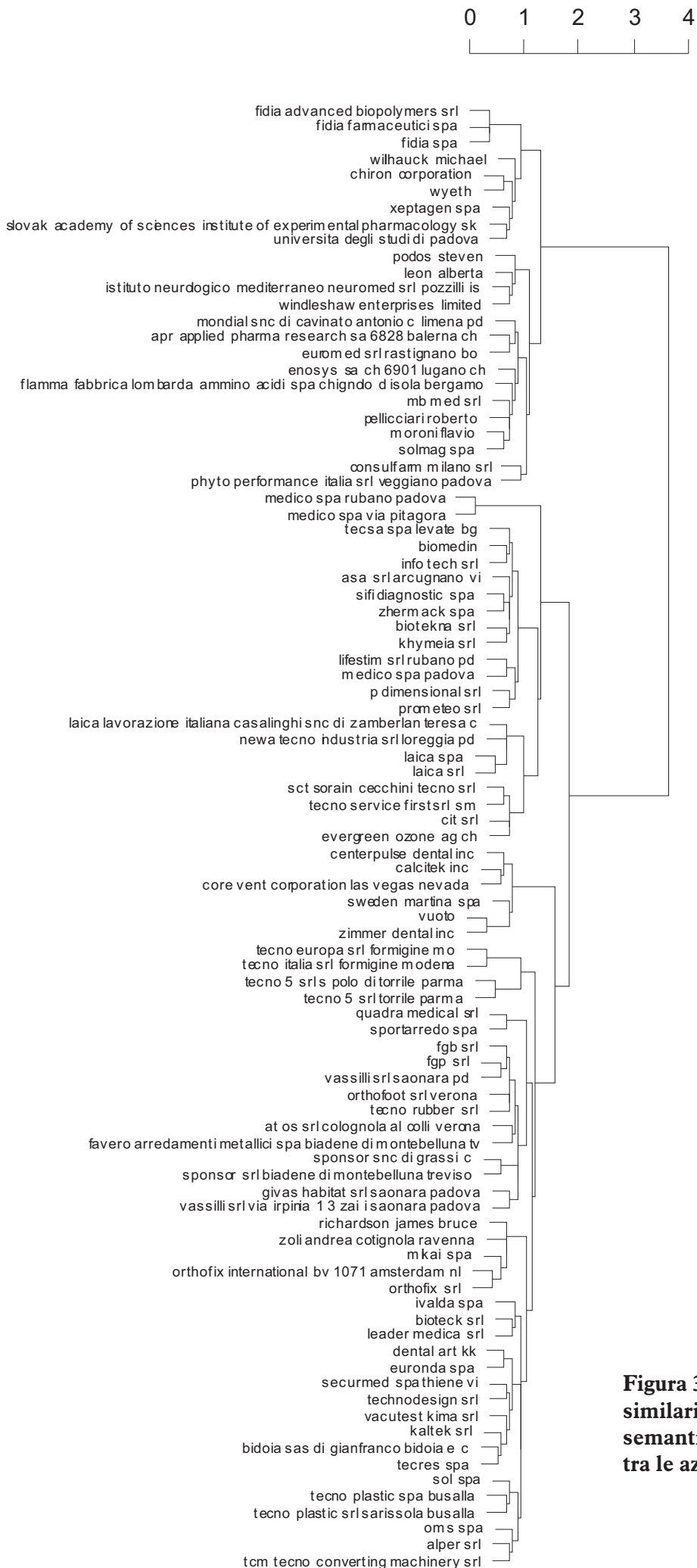
Figura 4 ci fa vedere come il biomedicale sia nettamente diviso in due parti distinte, identificabili da due rami di aziende. Il ramo di sinistra è relativo a chimica e industria farmaceutica di base e nanotecnologie. Il ramo di destra invece è relativo a altri settori, quali l'ortodonzia, i prodotti ortopedici e altri servizi medicali.

Ricordiamo che i brevetti non sono solamente dei testi a contenuto tecnologico ma contestualizzano molto le invenzioni, e vanno a cogliere dimensioni anche diverse dagli strumenti fisici citati. Spesso si parla di rapporto con il paziente, o del "vantaggio" derivante dall'invenzione in termini di comodità, facilità di uso, estetica, etc. In questo senso le aziende dell'or-

todonzia possono essere poste accanto a quelle ortopediche.

Raggruppamenti tecnologici

Settore biomedicale veneto



BeP srl www.innovazionebusiness.it

Figura 3:
similarità
semantica
tra le aziende

V Alcuni trend tecnologici

Il dataset brevettuale comprende date di deposito che vanno dal 1978 al 2010. Gli anni antecedenti al 1986 non presentano un numero sufficientemente alto di brevetti da fornire informazioni elaborabili. Dalla Tabella 1 e dalla Tabella 2 abbiamo

visto come i settori più nominati siano il farmaceutico di base e le applicazioni ortodontiche, perciò concentriamoci su questi due settori per individuare eventuali tecnologie "rampanti" o obsolescenti.

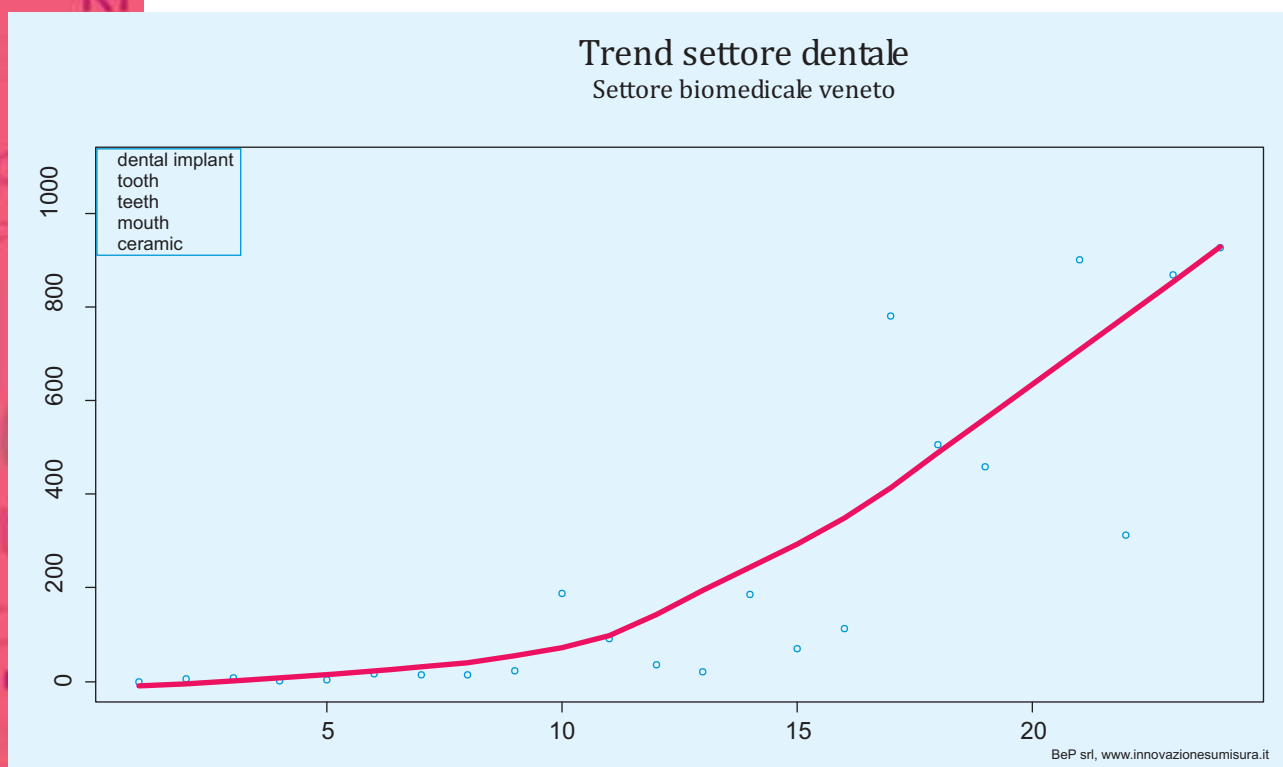


Figura 4: trend del settore dentale

Selezioniamo una serie di termini legati a questi settori; dalla Figura 4 e Figura 5 si può vedere come il settore dentario sia in crescita, mentre alcune tecnologie legate alla neurologia siano in calo.

Per analizzare al meglio l'andamento temporale al netto delle variazioni annuali del volume di brevetti, si è proceduto eseguendo:

1. standardizzazione per anno: le frequenze delle parole di ogni anno sono state

rapportate alla media e alla deviazione standard dell'anno;

2. stima di un modello lineare per ogni parola, per descrivere l'andamento nel tempo.

Questi due passi consentono di selezionare in modo automatico le parole che presentano un trend positivo o negativo, ovvero che presentano un coefficiente di regressione in calo o crescita. Il livello di significatività statistica è stata posta al 95%.

V.a Un altro esempio

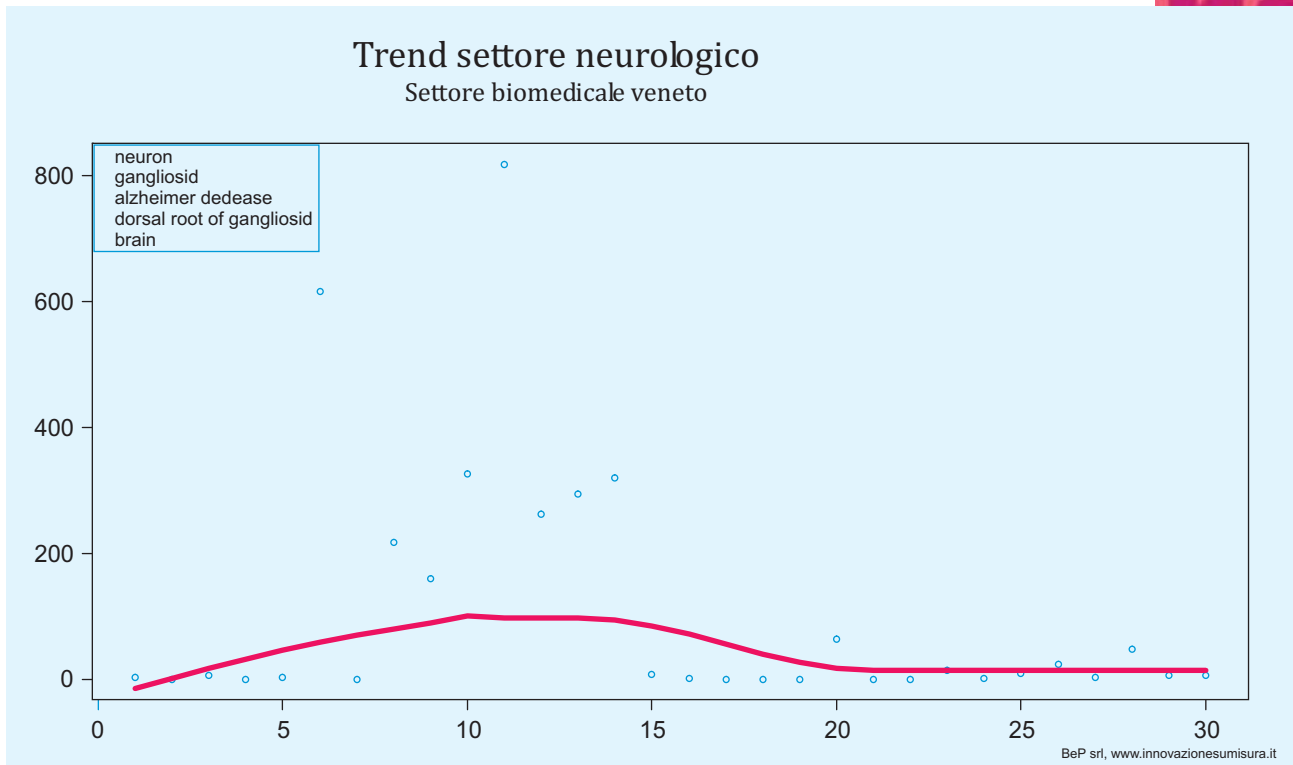


Figura 5: tecnologie in declino

VI Conclusioni e possibili sviluppi

Gli utilizzi futuri di queste metodologie sono vari: individuare le similarità tra aziende senza dover effettuare ricerche dispendiose in termini di tempo e risorse può portare a risparmi di tempo consistenti. Soprattutto considerando che le banche dati brevettuali utilizzate abbracciano l'intera letteratura brevettuale mondiale. C'è la possibilità di ottenere un occhio discreto sui trend tecnologici mondiali, suddividendo per categorie tecnologiche.

Possiamo immaginare un sistema di *information retrieval* che risponda alle nostre sollecitazioni in modo interattivo. Possiamo immaginare di far collocare un nostro prodotto all'interno del mondo conosciuto, oppure individuare i partner potenzialmente sinergici alla nostra azienda.

Finito di stampare nel mese dicembre 2011 a cura di Il Prato srl - Saonara (Padova)



Via Croce Rossa, 56 - 35129 Padova
Tel. 049.8062236 fax 049.8062200
e-mail: innovazione@pd.cna.it