

# Il valore del dato destrutturato (2a parte). Il contributo del dato web nella misurazione dello skill mismatch

*Mario Mezzanica, Roberto Boselli*

Università degli Studi di Milano Bicocca - CRISP

[mario.mezzanica@unimib.it](mailto:mario.mezzanica@unimib.it), [roberto.boselli@unimib.it](mailto:roberto.boselli@unimib.it)

## 1. Introduzione

Prosegue l'indagine sul valore del dato destrutturato iniziata nel numero scorso [1]. L'indagine si concentra su un'analisi del mercato del lavoro italiano svolta attraverso l'utilizzo di più di 170 mila annunci di offerte di lavoro pubblicati sul web. Gli annunci sono stati sottoposti ad un'analisi testuale e statistica per coglierne le specificità ed il valore quali e quantitativo.

In questo numero si cercherà di rispondere alla seguente domanda: quale contributo può aggiungere il dato web destrutturato alla conoscenza di un fenomeno complesso come il mercato del lavoro?

E' necessario trattare il dato destrutturato per portarlo ad un alto livello di qualità e di credibilità affinché possa restituire tutto il suo potenziale conoscitivo. Nel numero scorso si è descritta la metodologia di trattamento dei dati destrutturati estratti dal web, gli annunci di lavoro, per ottenere informazioni utili ad analisi statistiche di tipo quantitativo [1]. L'obiettivo di questa seconda parte della ricerca è verificare la significatività e specificità delle informazioni derivanti dai dati web, concentrando l'attenzione ora sugli aspetti qualitativi. In particolare, si cerca di evidenziare il valore aggiunto apportato da tali dati in termini di qualità e di profondità su alcuni temi specifici, le skills delle professioni maggiormente ricercate sul web. Tale valore aggiunto mostra come i dati del web possono essere utilizzati in modo complementare con i dati amministrativi, o derivanti da analisi statistiche, anche per affrontare lo skill mismatch, un problema ben noto del mercato del lavoro.

## 2. Il processo di recruitment online

Il web è sempre più usato da chi offre e chi cerca lavoro per la pubblicazione degli annunci e dei curricula. Le tecnologie informatiche permettono la creazione e diffusione di nuovi canali e di nuovi approcci all'intermediazione tra domanda e offerta, dando vita prima all'e-recruitment e poi, con l'avvento dei social media, al social recruitment [2][3]. Il processo di recruitment è caratterizzato dalla pubblicazione su un sito web, da parte di chi offre lavoro, di un annuncio di lavoro, di solito composto da informazioni di contatto e dalle informazioni relative alla posizione aperta. Chi cerca lavoro ha la possibilità di caricare il proprio cv, spesso riempiendo dei form predisposti dai siti, e poi inizia a cercare nei diversi siti le offerte che più si avvicinano ai propri desideri [4].

Annunci e cv possono essere pubblicati su diverse tipologie di siti di recruitment: i cosiddetti job boards e i motori di ricerca specializzati, come Monster o Career builder, che pubblicano le offerte di lavoro delle aziende; gli aggregatori, come Jobrapido e Indeed, che raccolgono gli annunci da diversi siti, inclusi i job boards, facilitando e riducendo i tempi di ricerca; i siti aziendali; i siti delle agenzie per il lavoro, e i social media. Questi ultimi facilitano lo stabilirsi di relazioni online e permettono ai candidati di pubblicare il proprio profilo professionale arricchito delle informazioni sulla propria rete di contatti. I selezionatori utilizzano sempre più le social networks, come LinkedIn, con l'obiettivo di conoscere maggiormente la rete di contatti dei candidati e per verificare la veridicità delle informazioni presentate nei curricula.

Il matching tra domanda e offerta avviene, o dovrebbe avvenire, sulla base della corrispondenza tra le caratteristiche del lavoro descritto nelle offerte e le caratteristiche del profilo del candidato. Tra queste ultime rivestono un ruolo fondamentale le competenze e le skill, che possono essere suddivise in skill personali, di base e professionali. Esse sono le informazioni più importanti per permettere il matching, tuttavia non sono ancora state utilizzate a fondo dai siti e sistemi di recruitment online<sup>1</sup>.

### 3. Il matching basato sulle skill

In tutte le tipologie di siti sopra descritti il punto di vista da cui parte il servizio è quello dell'azienda che espone la domanda aspettando che il candidato si offra per il lavoro. Inoltre questi siti presentano alcuni punti critici: processi lenti e inefficienti di selezione, poca rilevanza data alle skill, mancanza di trattamento della semantica delle informazioni, tutto ciò comporta una conseguente bassa integrazione tra domanda e offerta.

Un servizio innovativo ed efficace di intermediazione online dovrebbe quindi valorizzare prima di tutto il ruolo del candidato mettendolo al centro del processo di servizio, aumentare le capacità di selezione delle aziende, e soprattutto valorizzare il ruolo delle skill nel processo di matching.

Le skill possono essere ricavate sia dalla domanda (dai siti aggregatori) sia dall'offerta (social media e sistemi che permettono l'autovalutazione), e permettono il matching sulla base sia delle informazioni ricavabili dalle relazioni presenti nelle social network, sia dalle opportunità aperte e pubblicate sul web. E' necessario inoltre sviluppare sistemi di misurazione del ranking delle skill e del gap tra skill possedute e richieste.

Per sviluppare un sistema di identificazione delle skill è prima di tutto necessario creare un dizionario aggiornato (o sistema di classificazione) delle skill che permetta una diretta corrispondenza tra skill e professione (non esiste ancora in Europa un sistema paragonabile a O\*NET, utilizzato negli USA). Inoltre, sono necessarie tecniche di trattamento del testo e dei dati non strutturati e semi-strutturati per estrarre dagli annunci e dai curricula presenti sul web le skill e inserirle nel dizionario. Questo ultimo obiettivo ha guidato la parte dell'indagine

effettuata sugli annunci di lavoro estratti dal web che di seguito presentiamo.

### 4. La qualità del dato web e il problema delle skill

Uno dei contributi principali forniti dai dati web alla conoscenza del mercato del lavoro può essere l'identificazione delle skill richieste negli annunci di lavoro. Tale contributo può inoltre essere utilizzato per colmare il cosiddetto *skill mismatch*. Il mismatch è una caratteristica basilare del mercato del lavoro definendo quest'ultimo come il luogo dove i lavoratori possono trovare un match con un'occupazione. È quasi impossibile trovare un match perfetto per ogni lavoratore, e dunque la dimensione e le caratteristiche del mismatch possono costituire un indicatore decisivo del funzionamento del mercato del lavoro. Dal punto di vista generale la letteratura distingue due tipologie di mismatch: un mismatch quantitativo e uno qualitativo. Il primo è relativo alla differenza tra il numero di posti domandati e offerti, il secondo è relativo alle caratteristiche della domanda di lavoro rispetto all'offerta. In altri termini un mismatch qualitativo è ascrivibile alla differenza tra le skill e le competenze del lavoratore rispetto a quelle che sono richieste dal lavoro.

Tale mismatch può essere causato da diversi fattori [5]. Durante periodi di recessione, per esempio, il tasso di distruzione dei posti di lavoro supera il tasso di creazione determinando l'aumento della disoccupazione. In presenza di tassi di disoccupazione elevati i lavoratori qualificati possono decidere di accettare offerte di lavoro che richiedono qualifiche o competenze inferiori anziché attendere un match più adeguato. Tutto ciò genera *overqualification* o *overskilling*. Allo stesso modo durante periodi di crescita elevata si può registrare il fenomeno opposto: le imprese, trovando difficoltà a riempire le *vacancies* aperte con le persone dalle competenze richieste, possono decidere di offrire il lavoro a chi non possiede del tutto le competenze necessarie anziché attendere di trovare un match corretto. In questo caso si parla di *underqualification*.

Il match ottimale tra occupazione e lavoratore si ha quando una certa combinazione di skill richieste dal lavoro si riflette in quelle offerte dal lavoratore. E' necessaria quindi una misura del mismatch che possa misurare tutte le componenti,

<sup>1</sup> Un esempio che va in questa direzione è il sito americano [myskillsmyfuture.org](http://myskillsmyfuture.org)

oggettive e soggettive, formali e informali, delle competenze. Tuttavia non esiste ancora una misura adeguata per il mismatch tra skill. Le cause risiedono nella natura qualitativa del mismatch che si riferisce alle caratteristiche più complessive del lavoratore e dell'occupazione.

Le misure esistenti si basano principalmente su qualche forma di classificazione e standardizzazione. Per esempio, per quanto riguarda le occupazioni si possono usare la classificazione internazionale ISCO (International Standard Classification of Occupations) o l'italiana CP2011 sviluppata da ISTAT, mentre per le skill si può utilizzare il sistema O\*NET (Occupational Information Network) recepito in Italia da Isfol. Tali sistemi però non riescono a catturare la complessità delle skill richieste da ciascuna occupazione. Inoltre non esistono variabili economiche che possano permettere una quantificazione precisa delle skill.

La soluzione maggiormente utilizzata è quindi quella di eseguire delle *skill survey*, somministrare questionari e indagini a campioni di lavoratori. In Italia dal 2011 Excelsior si occupa di rilevare le competenze trasversali con questo tipo di indagine. Più recentemente, su scala europea, il Cedefop sta implementando la prima *skill survey* europea somministrata a tutti i paesi membri contemporaneamente e con una metodologia comune.

Tuttavia tali indagini hanno dei limiti, tre in particolare: 1) sono molto costose, sia considerando i costi diretti (l'implementazione) sia indiretti (rapporto costo-opportunità); 2) la loro implementazione ed elaborazione sono complesse, generalmente effettuate a scadenza annuale o biennale; 3) l'approccio è di tipo *top-down* con risposte ed elenchi di skill predefiniti.

In tale contesto, l'analisi dei dati del web potrebbero aiutare a superare i limiti sopra descritti delle *skill survey*. I dati qualitativi possono infatti offrire un valido strumento complementare rispetto alle *skill survey* nella misurazione di skills e competenze. In primo luogo essi superano i limiti dei sistemi fissi di classificazione sfruttando un approccio opposto, *bottom-up*, e catturando l'evoluzione del linguaggio e della terminologia (le descrizioni cambiano nel tempo a seconda dei requisiti delle aziende e del mercato).

Secondo, i dati web permettono di identificare le skill emergenti nel mercato tramite le offerte presenti nel web. Terzo, l'analisi di tali dati ha costi minori e si basa su dati in tempo reale, quindi sempre aggiornati.

## 5. Analisi delle figure emergenti nel web

Grazie all'analisi eseguita sugli annunci di lavoro del web abbiamo potuto identificare alcune figure professionali emergenti, le più richieste con le skill desiderate dalle aziende. Seguendo la metodologia descritta in [1] è stato possibile iniziare a creare un dizionario delle skill che può costituire una fonte preziosa di informazione per una molteplicità di soggetti. Le skill sono state divise in tre macro-gruppi:

- Skill legate alla persona: competenze ed attitudini di carattere trasversale legate a tratti della personalità e di atteggiamento caratteristici dell'individuo;
- Skill legate alla professione: competenze tipiche e specifiche della professione, trasmesse prevalentemente tramite l'esperienza stessa del lavoro o attraverso canali di formazione di carattere professionale;
- Conoscenze di base: abilità standardizzate, spesso trasversali tra le occupazioni, acquisite generalmente attraverso canali educativi formali (scuola, università).

Inoltre, le figure emergenti possono essere suddivise in alto, medio o basso livello di istruzione. Tale suddivisione si basa sulla presenza negli annunci dei riferimenti ai titoli di studio richiesti da una determinata occupazione.

Un esempio può chiarire meglio quanto detto. Se paragoniamo la descrizione di un'occupazione fornita da un sistema di classificazione ufficiale (per esempio CP2011) con quella estratta dal web, avremo nella prima una più ricca descrizione delle attività che competono a quell'occupazione, mentre nella seconda avremo molte più informazioni relative agli aspetti qualitativi legati alla professione.

In figura 1 è riportata un'offerta di lavoro sul web per la professione di ingegnere ambientale.

<body> per prestigiosa azienda cliente, 1 Ingegnere Ambientale. Fondamentale Laurea in Ingegneria ambientale. Indispensabile utilizzo Autocad 2D. Minima conoscenza dell'inglese. Conoscenza del SGA SGSL e elaborazioni manualistiche. Automuniti. Zona di lavoro vicinanze Casale Monferrato.</body>

Figura 1. Esempio di offerta di lavoro sul web.

La classificazione ufficiale delle professioni

CP2011 definisce gli ingegneri ambientali come segue:

#### 2.2.1.6.1 - Ingegneri edili e ambientali

*Le professioni comprese in questa unità conducono ricerche ovvero applicano le conoscenze esistenti nel campo della pianificazione urbana e del territorio, della progettazione, della costruzione e della manutenzione di edifici, strade, ferrovie, aeroporti, ponti e sistemi per lo smaltimento dei rifiuti e di altre costruzioni civili e industriali. Definiscono e progettano standard e procedure per garantire la funzionalità e la sicurezza delle strutture. Progettano soluzioni per prevenire, controllare o risanare gli impatti negativi dell'attività antropica sull'ambiente; conducono valutazioni di impatto ambientale di progetti ed opere dell'ingegneria civile o di altre attività; si occupano di prevenzione e risanamento dei fenomeni di dissesto idrogeologico e instabilità dei versanti, di sistemazione e gestione dei bacini idrografici. Sovrintendono e dirigono tali attività.*

Come si può osservare le due fonti (web e canale ufficiale) forniscono informazioni complementari tra loro, che consentono di avere una visione completa degli aspetti qualitativi legati alla professione. Attraverso la costruzione di una tassonomia apposita, con i metodi descritti in [6], vengono estratti dall'esempio del web le skills associate e classificate nel modo seguente: come skill di base la conoscenza della lingua inglese e la patente, come skill legate alla professione

l'utilizzo di software e la conoscenza di modelli specifici (Autocad 2D e SGA SGSL), mentre non sono presenti skill legate alla persona.

La rappresentazione grafica delle skill (figura 2) può essere fornita attraverso un word-cloud che presenta una dimensione del testo direttamente proporzionale al numero di volte in cui ciascuna skill viene richiesta nelle offerte di lavoro per ciascuna professione.

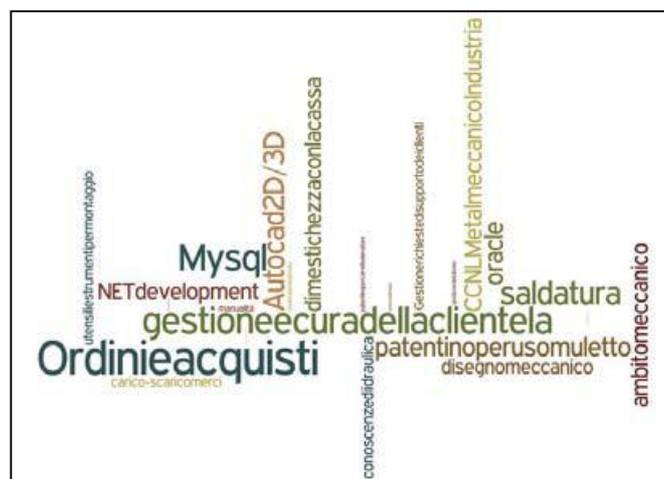


Figura 2. Word cloud skills della professione, tutte le professioni.

Analizzando tutti gli annunci scaricati abbiamo potuto quantificare la distribuzione percentuale delle diverse skill. Si è notato che le principali skill personali richieste sono di carattere relazionale (20% del totale degli annunci che richiedono skill personali), ma anche skill che hanno a che fare con l'autonomia lavorativa (15%) e le capacità di risolvere i problemi (14%). In particolare, le capacità relazionali e soprattutto le capacità di gestire la complessità (problem solving, lavorare per obiettivi, etc.) sono molto importanti tra le professioni high level e tra quelle medium level non manuali. Passando alle competenze di base, predomina su tutte la conoscenza dell'inglese (58%) e la conoscenza del pacchetto Office (24%). Per quanto riguarda le competenze specifiche della professione non è possibile trarre indicazioni precise stante l'elevata eterogeneità delle informazioni contenute. Tuttavia la figura 2 mostra le potenzialità dell'analisi delle domande di lavoro sul web, dalle quali si può rilevare richieste prevalenti di alcune competenze specifiche (per esempio il programma MySQL).

## 6. Conclusioni

Il dato qualitativo che emerge dalle domande di lavoro postate sul web permette di risolvere parzialmente il problema del mismatch delle skill perché fornisce una descrizione più ricca degli aspetti qualitativi delle professioni. La maggiore diffusione e profondità delle informazioni consente quindi un'ampliamento della potenziale platea di possibili destinatari contribuendo a ridurre di conseguenza il mismatch. Chi cerca lavoro può comprendere meglio quale tipo di skill viene richiesta per la professione ricercata, le aziende possono rendere più efficace la ricerca del candidato ideale, e chi si occupa di formazione può indirizzare il percorso di apprendimento in modo più mirato.

Tuttavia, occorre sottolineare come, nonostante i vantaggi descritti sopra, i dati web devono essere visti come complementari e non sostitutivi rispetto agli altri approcci di rilevazione delle competenze. Soprattutto a causa delle seguenti criticità dei dati

web. In primo luogo le domande di lavoro postate sul web costituiscono solo una parte delle domande di lavoro complessive. In particolare, le tipologie di occupazioni maggiormente richieste tramite il web non corrispondono alla totalità delle figure professionali presenti. Inoltre, molte imprese preferiscono reclutare le figure apicali o quelle ad elevata specializzazione, attraverso agenzie dedicate o *head hunters* specializzati, piuttosto che affidarsi ad annunci pubblici su internet. Infine la composizione settoriale delle domande web mostra una certa sovrarappresentazione delle figure professionali specializzate nell'industria rispetto a quelle dei servizi.

## Bibliografia

- [1] Mezzanzanica, M., Boselli, R., Il valore del dato destrutturato: Una metodologia di analisi delle offerte di lavoro sul web (1a parte), *Statistica e Società*, Anno 2, N.3, 2013, pp. 5-8.
- [2] Lee I., The evolution of e-recruiting: a content analysis of Fortune 100 career web sites, *Journal of Electronic Commerce in Organizations*, 3 (3), pp. 57-68, 2005.
- [3] Boselli, R., Cesarini, M., Mercurio, F. e Mezzanzanica, M. (2014) How the Social Media Contributes to the Recruitment Process?, *Proceedings of European Conference on Social Media (ECSM 2014)*, 10-11 July, Brighton, UK.
- [4] Singh, A., Rose, C., Visweswariah, K., Chenthamarakshan, V. and Kambhatla, N. (2010) "PROSPECT: a System for Screening Candidates for Recruitment", *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 659-668.
- [5] Cedefop (2010), The skill matching challenge; analysing skill mismatch and policy Implications.
- [6] Boselli, R., Colombo, E., Graziani, C., Mezzanzanica, M., Ronzoni, G., L'offerta di lavoro sul web, in Mezzanzanica, M., (a cura di) *Ipotesi di lavoro. Le dinamiche, i servizi e i giudizi che cambiano il mercato del lavoro*, Aracne, Roma, ottobre 2013, ISBN: 978-88-548-6541-9, pp. 77-134.