



Strat our

Lavoro Interactive 3.0.

A cura di

**Vincenzo Lombardo (Responsabile scientifico),
Rossana Damiano, Cristina Gena**

Report dell'attività di progetto

22 ottobre 2013





PREMESSA.....	4
Metodologia e timeline del progetto.....	8
Modello generale, architettura del sistema e flusso di interazione utente/sistema	9
Basi di conoscenza utilizzate dal sistema	11
Descrizione delle professioni.....	11
Conoscenza semantico-lessicale.....	12
Modello parametrico	13
SUGGERIMENTO DI TAG.....	18
I tag nel dataset di IoLavoro (ottobre 2012).....	19
Principi del suggerimento di tag	22
Il suggerimento di tag in Lavoro Interactive 3.0	23
A. Indice di similarità tra i tag.....	24
B. Indice di similarità tra profili e tag.....	26
C. Indice di similarità tra profili.....	27
Valutazione e scelta della strategia finale di suggerimento di tag.....	29
Raffinamento della strategia di suggerimento di tag	31
Altre tecniche statistiche	31
Profili professionali nella classificazione di IoLavoro	32
Interazione tra utenti candidato e utenti azienda tramite il sistema.....	33
MultiWordNet.....	33
PROCEDURA DI MATCHMAKING.....	36
Strutturazione degli annunci	37
Modello utente	38
Similarità tra annunci e candidati.....	40
Match tra tag.....	40
Definizioni	41
Match lessicale	41
Algoritmo per match lessicale:.....	43
Match semantico	43
Criteri di assegnazione dei valori ai tipi di match.....	44
Valori assegnati.....	45
Applicazione del match lessicale e semantico agli altri campi della descrizione dell’annuncio:	45
Assegnamento dei pesi	46
Esempi	47
Applicazione del Case Based Reasoning (CBR).....	49
Strategie di CBR e presentazione dei risultati	52
Lato azienda	52
Formule di matchmaking (suggerimenti) - prima versione.....	52
Suggerimento ristretto a1 - prima versione.....	52
Suggerimento con ampliamento della ricerca a2 – prima versione.....	53
Suggerimento stupiscimi a3 - prima versione.....	53
Lato candidati	54
Suggerimento di default c1 - prima versione	54
Suggerimento con filtraggio c2 – prima versione	54
Suggerimento stupiscimi c3 – prima versione	55
Test effettuati con il primo modello valorizzato	55
Suggerimenti seconda versione (post prima valutazione)	59
Suggerimento ristretto A1 - seconda versione	59
Suggerimento con ampliamento della ricerca A2 – seconda versione.....	61
Suggerimento stupiscimi A3 - seconda versione	62
Lato candidati	63
Suggerimento di default C1 - seconda versione	63
Suggerimento con filtraggio C2 – seconda versione (= prima versione).....	64



Suggerimento stupiscimi C3 – seconda versione.....	65
Test effettuati con il secondo modello valorizzato.....	65
Presentazione dei risultati e test sul campo (Edizione IoLavoro di Ottobre 2013).....	68
CONCLUSIONI E PROSPETTIVE FUTURE	70
RELEASE E CARATTERISTICHE TECNICHE RICHIESTE PER IL SERVER.....	73
RIFERIMENTI BIBLIOGRAFICI SELEZIONATI	74
APPENDICE: PRESENTAZIONE DEI RISULTATI.....	76
Portale Lavoro Interactive 3.0 - Home.....	76
Lato aziende – Benvenuto nella pagina personale aziende	77
Lato aziende – Gestione propri annunci	77
Lato aziende – Gestione di un annuncio.....	78
Lato aziende – Visualizzazione autocandidature con grado di compatibilità	78
Lato aziende – Visualizzazione candidati del sistema per un annuncio (A1).....	79
Lato aziende – Visualizzazione candidati del sistema per un annuncio (A2).....	79
Lato aziende – Visualizzazione candidati del sistema per un annuncio (A3).....	80
Lato candidati – Pagina di benvenuto	80
Lato candidati – Visualizzazione annunci per candidato (C1)	81
Lato candidati – Visualizzazione annunci per candidato per profilo (C2).....	81
Lato candidati – Visualizzazione annunci per candidato C3	82
Fiera virtuale 3D – Home per utente non loggato.....	82
Fiera virtuale 3D – Ricerca libera	83
Fiera virtuale 3D – Risultati ricerca libera.....	83
Fiera virtuale 3D – Home – utente loggato (A1).....	84

La ricerca di lavoro online è oggi molto diffusa e sono numerosi i siti commerciali che appartengono a questa tipologia (Figura 1)¹. Tali siti permettono agli utenti di effettuare ricerche in un *database* di offerte di lavoro e offrono la possibilità, agli utenti registrati, di candidarsi per una determinata offerta (o, nel caso di un datore di lavoro, di contattare i candidati selezionati dati i curricula inseriti). Nonostante la varietà dell'offerta, tuttavia, la ricerca di lavoro online non è, a tutt'oggi, un compito semplice, come appare evidente da una semplice comparazione dei siti suddetti, che differiscono per funzionalità di ricerca, descrizione di annunci e candidati, utilizzo dei tag, ecc. .

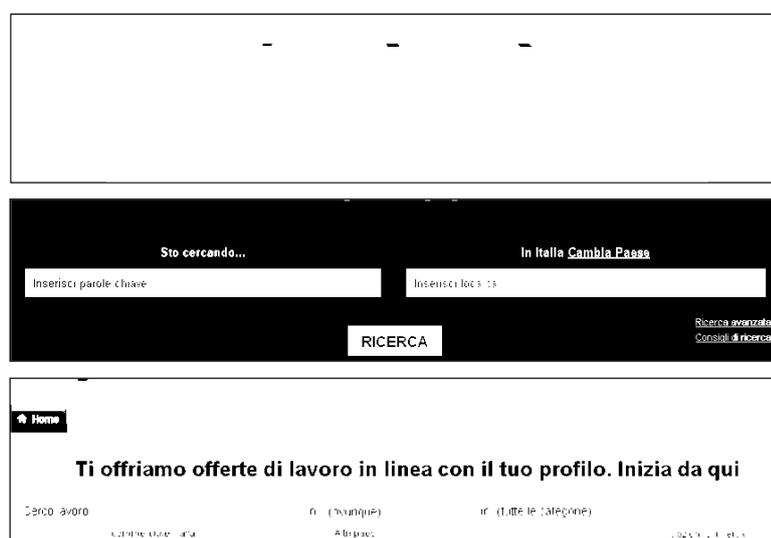


Figura 1: Alcune maschere di ricerca di portali specializzati nella ricerca di lavoro online. Dall'alto: Monster (www.monster.it), HelpLavoro (www.helplavoro.it), InfoJobs (www.infojobs.it). Profilo e luogo sono predominanti, ma ogni sito presenta un approccio diverso (ad esempio, solo nella maschera in basso compare il concetto di categoria).

Uno degli inconvenienti principali della ricerca di lavoro online è dato dal fatto che ogni sito categorizza le offerte di lavoro secondo una classificazione diversa: ciò richiede all'aspirante lavoratore uno sforzo considerevole per collocarsi all'interno della classificazione di ogni sito. Alcuni siti mettono a disposizione degli utenti una funzionalità di ricerca avanzata per rendere più efficace e completa la ricerca nelle offerte di lavoro, ma questo strumento è reso meno efficace dal fatto che lo schema complessivo (Figura 2) secondo cui viene descritto l'annuncio (e sul quale si basa la ricerca avanzata) cambia da sito a sito, con campi il cui significato non è sempre noto all'utente.

¹ www.monster.it, www.infojobs.it, www.miojob.repubblica.it per citare solo alcuni dei più noti.

Riepilogo Annuncio			
Azienda Kinet Human Resources S.r.l.	Descrizione Annuncio Ricerca anno per azienda cliente Uno SVILUPPATORE JAVA/JSP-Pagina Terzo Nord.	Pubblicato il	01/11/2012
Località Torino (Ormai), Piemonte Italia	Requisiti richiesti: Esperienza maturata nella programmazione con Java e JSP, eventuale merito C++ e PHP/MySQL Conoscenza di Database relazionali; Conoscenza del principale framework di sviluppo; Buona conoscenza della lingua inglese sia compreso, scritto e parlato, gradita a conoscenza di altre lingue. Flessibilità e volontà nel apprendere merito; Capacità di problem solving; Disponibilità a trasferirsi in Italia e all'estero.	Modalità di candidatura	Autoregolata (Candidatura)
Settore Agenzie di ricerca personale/Staffing/Agenzie di recruiting	Versione da sviluppare: Attività di programmazione front-end in applicazioni Web. Zona di lavoro: Torino Nord Orario Full-time	[Mostra altre informazioni]	
Contratto Full Time	Modalità di contratto: Attività di programmazione front-end in applicazioni Web. Zona di lavoro: Torino Nord Orario Full-time	Luogo	
Anni Esperienza Da 5 a 7 anni	Modalità di contratto: Attività di programmazione front-end in applicazioni Web. Zona di lavoro: Torino Nord Orario Full-time	Città	Ormai
Livello Istruzione Diploma	Modalità di contratto: Attività di programmazione front-end in applicazioni Web. Zona di lavoro: Torino Nord Orario Full-time	Indirizzo	via ...
Livello di Carriera Con esperienza	Modalità di contratto: Attività di programmazione front-end in applicazioni Web. Zona di lavoro: Torino Nord Orario Full-time	Capozone	Italia
		Descrizione	
		Titolazione/Qualifica	Sviluppatore (DB, Java, ...)
		Categorie	Area ...
		Settore	.../...
		Numero di requisiti	1/2
		Numero di candidati	...

Figura 2: Esempio di descrizione di posizione lavorativa in alcuni siti commerciali. Si notino le differenze nei campi descrittivi, categorie, livello di dettaglio, ecc.

In conclusione, la ricerca per area geografica e per titolo di studio sono di fatto predominanti rispetto a una caratterizzazione a 360° del candidato e dell'azienda, con il rischio che molti candidati, scoraggiati dalla difficoltà di trovare offerte di lavoro interessanti, si limitino a verificare con cadenza regolare le nuove offerte di lavoro nella home page del sito (cosiddetto criterio di *recentezza*), magari filtrate per area geografica o per titolo di studio.

In parallelo con lo sviluppo dei siti commerciali, alcune iniziative di ricerca hanno posto l'accento sulla descrizione dei candidati, riconoscendo il fatto che criteri espliciti e non ambigui per la descrizione di offerte di lavoro e candidati costituiscono una premessa necessaria all'incontro tra domanda e offerta di lavoro. Per esempio, i progetti TIPTOE e LO-MATCH² hanno sviluppato strumenti automatici che aiutano i candidati a descrivere il proprio curriculum secondo uno schema standardizzato e non ambiguo delle competenze professionali, al fine di supportare il *matchmaking* (corrispondenza, incrocio) automatico tra curricula e offerte di lavoro. Tali progetti, tuttavia, si basano su iniziative europee per la descrizione delle competenze secondo il *framework* standardizzato EQF-EQVET³, tuttora in fase di consolidamento e non ancora disponibili nello scenario italiano.

² Gatteschi V., Lamberti F., Demartini C., *On the Use of Semantic Technologies to Support Education, Mobility and Employability*, Semantic Hyper/Multimedia Adaptation, pp. 127-150, 2012, Springer;
Gatteschi V., Lamberti F., Demartini C., van Wezel R., Bettiol S., *Exploiting semantics for constructing and comparing occupational and educational-driven qualifications: the TIPTOE project*, Journal of Universal Computer Science, vol. 18, n.1, 2012.

³ European Qualification Framework (EQF, disponibile al sito http://ec.europa.eu/eqf/home_en.htm) è il framework di riferimento per la comparazione dei sistemi nazionali di qualifica professionale; European Credit System for Vocational Education and Training (EQVET, <http://www.ecvet-team.eu>) è il framework europeo per la creditizzazione della formazione professionale.

Lavoro Interactive si propone di affrontare il problema con un approccio centrato sugli utenti, adattando metodologie dall'intelligenza artificiale e dall'analisi statistica dei dati.

Il portale Lavoro Interactive 3.0 è una piattaforma innovativa per la ricerca di lavoro online caratterizzata da un approccio 3.0"; è stata realizzata, nell'ambito dell'azione 3 del progetto strategico STRATTOUR (Programma Alcotra 2007-2013), dal Centro Interdipartimentale di ricerca sulla multimedialità e l'audiovisivo dell'Università degli studi di Torino (CIRMA), attraverso una convenzione stipulata il 25 ottobre 2012 tra Agenzia Piemonte Lavoro (APL) - soggetto attuatore della Regione Piemonte/partner, Agenzia Liguria Lavoro (ALL) - soggetto attuatore della Regione Liguria/capofila, Regione Valle d'Aosta/partner (VdA) ed il CIRMA stesso. Nel Web odierno, l'espressione "3.0" indica un sito in grado di assistere l'utente in maniera adattiva nel raggiungimento delle sue finalità, attraverso la predisposizione di un ambiente in cui l'utente viene coinvolto in un'interazione collaborativa e spontanea con il sistema, tramite modalità mutate dal mondo reale. Nel caso di Lavoro Interactive 3.0, l'ambiente di lavoro include:

- un portale informativo e interattivo per la ricerca di lavoro online, che – oltre alle funzionalità tipiche della ricerca di lavoro online – riconosce la necessità di supportare il candidato fin dalla fase di descrizione del proprio curriculum, con strumenti specifici per supportare l'incontro domanda-offerta (Fig. 3);
- una fiera virtuale del lavoro, realizzata attraverso l'utilizzo della tecnologia 3D, che rende l'interazione tipica della fiera reale, con modalità più naturali e coinvolgenti rispetto alla modalità ipertestuale.



Figura 3: Il problema dell'incontro domanda - offerta di lavoro, visto come associazione tra candidati e annunci (matchmaking, candidati a sx e annunci a dx).

Le funzionalità innovative di Lavoro Interactive 3.0 sono di tipo *proattivo* (cioè si attivano sulla base dei soli dati presenti nel sistema, in alcuni passaggi dell'interazione con l'utente, senza bisogno di un input esplicito dal parte dell'utente) e si basano su tecnologie di intelligenza artificiale messe a punto nel corso dell'ultimo decennio dalla ricerca sulle piattaforme sociali e semantiche nel Web: in particolare, Lavoro Interactive 3.0 utilizza tecniche di modellazione dell'utente (*user modeling*) per rappresentare annunci e candidati in maniera strutturata e si basa su tecniche statistiche e semantiche (tra cui *collaborative filtering* e *case-based reasoning*) per supportare l'utente (candidato o azienda) nell'inserimento dei dati e nella ricerca di annunci (o candidati).

In sintesi, le funzionalità innovative di Lavoro Interactive 2.0 sono le seguenti:

- **Suggerimento di tag** (*tag recommendation*): data l'importanza della descrizione di candidati e offerte di lavoro al fine dell'incontro tra domanda e offerta, il *tag recommender* suggerisce ai candidati e alle aziende *tag* appropriati, rispettivamente, al loro profilo professionale e ai loro annunci. Per fare ciò, il *tag recommender* si avvale di tecniche di tipo statistico che costituiscono lo stato dell'arte della ricerca nel settore dei *web recommender system* (sistemi di suggerimento per il web).
- **Match tra offerte di lavoro e candidati** (*matchmaking*): il *matchmaker*, superando i limiti dei filtri in un data base, suggerisce all'utente le offerte di lavoro potenzialmente rilevanti per lui/lei, e alle aziende i candidati maggiormente idonei, sulla base di un raffronto tra candidati e annunci basato sul paradigma summenzionato del *case based reasoning*. Quest'ultimo misura la similarità tra annunci e candidati secondo criteri emersi dalla collaborazione con APL, ALL e VdA e raffinati nel corso di un processo di valutazione semestrale condotto su dati reali.
- **Ricerca nelle offerte di lavoro**: la ricerca nelle offerte di lavoro, sebbene si presenti agli utenti come funzionalità a se stante, dipende logicamente dal *matchmaking*: mentre nel *matchmaking* propriamente detto tutti gli annunci presenti nel sistema vengono ordinati per rilevanza rispetto all'utente, nella ricerca nelle offerte di lavoro la rilevanza viene utilizzata per ordinare solo un sottoinsieme degli annunci, cioè gli annunci che rispettano i parametri di ricerca immessi dall'utente. Questa funzionalità è stata sviluppata solo per gli i candidati.

Il presente documento descrive in dettaglio le strategie che realizzano le funzionalità intelligenti di Lavoro Interactive. La proposta descritta è stata sviluppata a partire dall'analisi del *dataset* raccolto durante l'edizione di IoLavoro di ottobre 2012 e testata sull'edizione di IoLavoro di marzo 2013; l'implementazione del modello è stata disponibile per l'edizione di ottobre 2013 di IoLavoro, i cui risultati saranno analizzati successivamente alla chiusura di questo progetto. Il modello concettuale di descrizione dell'offerta e della domanda di lavoro e le strategie che ne supportano l'incontro riflettono quindi la struttura e il contenuto di tale *dataset*. L'utilizzo di paradigmi facilmente scalabili, quali il *case based reasoning*, tuttavia, li rende aperti alla possibilità di successivi aggiustamenti per adattarli a contesti diversi, nell'ambito dell'incontro tra domanda e offerta di lavoro.

Le funzionalità innovative di Lavoro Interactive sono state sviluppate secondo una metodologia di progettazione basata sui dati: in pratica, dati gli obiettivi del progetto (facilitare l'incontro tra domanda e offerta di lavoro attraverso un insieme di funzionalità), sono state individuati gli approcci maggiormente idonei a raggiungerli. In seguito, attraverso l'analisi dei dataset disponibili, tali approcci sono stati adattati al contesto specifico della fiera del lavoro, esemplificata da IoLavoro (nelle edizioni di ottobre 2012 e marzo 2013).

La metodologia adottata in Lavoro Interactive consiste di tre fasi (Figura 4):

1. la definizione di un modello generale, basato sull'esperienza precedente di CIRMA sui sistemi di raccomandazione: in questo modello si definiscono gli elementi principali che contribuiscono alla soluzione del problema, con le componenti di intelligenza artificiale e di analisi statistica; non contiene ancora i parametri effettivi che sono di interesse per il problema del *matchmaking* tra domanda e offerta di lavoro
2. la parametrizzazione del modello, basata sull'interazione con gli esperti: dagli incontri e le interviste con le agenzie regionali del lavoro coinvolte e dall'analisi della documentazione, soprattutto a supporto della manifestazione di IoLavoro, sono stati identificate le caratteristiche di interesse che si possono ritrovare nei dati (profili dei candidati e annunci di lavoro), suddivise tra obbligatorie e preferenziali, nonché le informazioni libere da tradurre in elementi del modello;
3. la valorizzazione dei parametri, basata sull'analisi dei database della manifestazione IoLavoro: il database di IoLavoro è diventato il *dataset* per il *tuning* del sistema, che ha guidato l'implementazione del sistema, da integrare nel portale; particolare attenzione è stata dedicata a una generalizzazione guidata del tuning, in modo da evitare problemi prevedibili di *overfitting* (cioè il problema di far dipendere eccessivamente dai dati di test il comportamento del sistema).

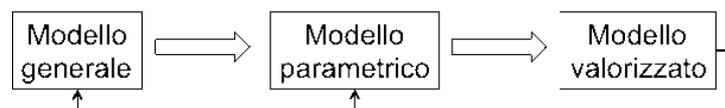


Figura 4: Metodologia seguita nel progetto Lavoro Interactive.

Lo sviluppo del progetto ha quindi seguito il cronoprogramma illustrato in Figura 5, per circa un anno e mezzo, ha utilizzato l'esperienza mutuata dalle edizioni in corso di IoLavoro, orientandone lo sviluppo futuro sulla base delle esperienze maturate in Lavoro Interactive, nell'ottica della fiera virtuale permanente. In particolare, un sforzo è stato dedicato all'estensione delle caratteristiche (*feature*) descrittive, rispetto alle precedenti edizioni di IoLavoro.

Maggio – Ottobre 2012	Elaborazione del modello generale, progettazione del sistema globale, definizione delle funzionalità e schematizzazione dell'architettura di riferimento.
Novembre – Dicembre 2012	Analisi <i>tagset</i> IoLavoro (ottobre 2012), progettazione e sviluppo del sistema parametrico di <i>tag recommendation</i>
Gennaio 2013	Valorizzazione del sistema parametrico di <i>tag recommendation</i> sul <i>tagset</i> di IoLavoro (ottobre 2012)
Febbraio – Marzo 2013	Design e sviluppo del modello di matchmaking
Aprile – Maggio 2013	Ottimizzazione matchmaking sulla valorizzazione fornita dal dataset di IoLavoro (marzo 2013)
Giugno – Settembre 2013	Implementazione, test e debug del motore di matchmaking
Ottobre 2013	Test sul campo per l'edizione di IoLavoro

Figura 5: Timeline della progettazione e sviluppo delle funzionalità innovative di Lavoro interactive descritte nel presente report.

L'architettura di Lavoro Interactive che implementa il modello generale è suddivisa in tre macro-moduli principali:

- Il **Web server** ospita il cuore del sistema, cioè il CMS, la gestione degli utenti, dei contenuti e della relazione tra di essi (cioè la navigazione);
- Il **Server di matchmaking** realizza le funzionalità di suggerimento di tag (*tag recommender*) e associazione automatica tra domanda e offerta di lavoro (*matchmaking*);
- I **Browser3D** e **HTML** realizzano, rispettivamente, la fruizione dei contenuti nella modalità "fiera 3D" e nella modalità ipertestuale standard.

All'interno del Server di Matchmaking, le funzionalità che esso realizza (*tag recommendation* e *matchmaking* propriamente detto) si basano sui dati contenuti nel **modello utente** e sulle sorgenti di conoscenza esterne, quali **database lessicali** (MultiWordNet) e **tassonomie** di dominio (tassonomia dei profili professionali). Il **modello utente** consiste in una rappresentazione strutturata degli utenti, che ne descrive caratteristiche e preferenze, sia attraverso i dati forniti dagli utenti, sia attraverso l'interazione con il sistema.

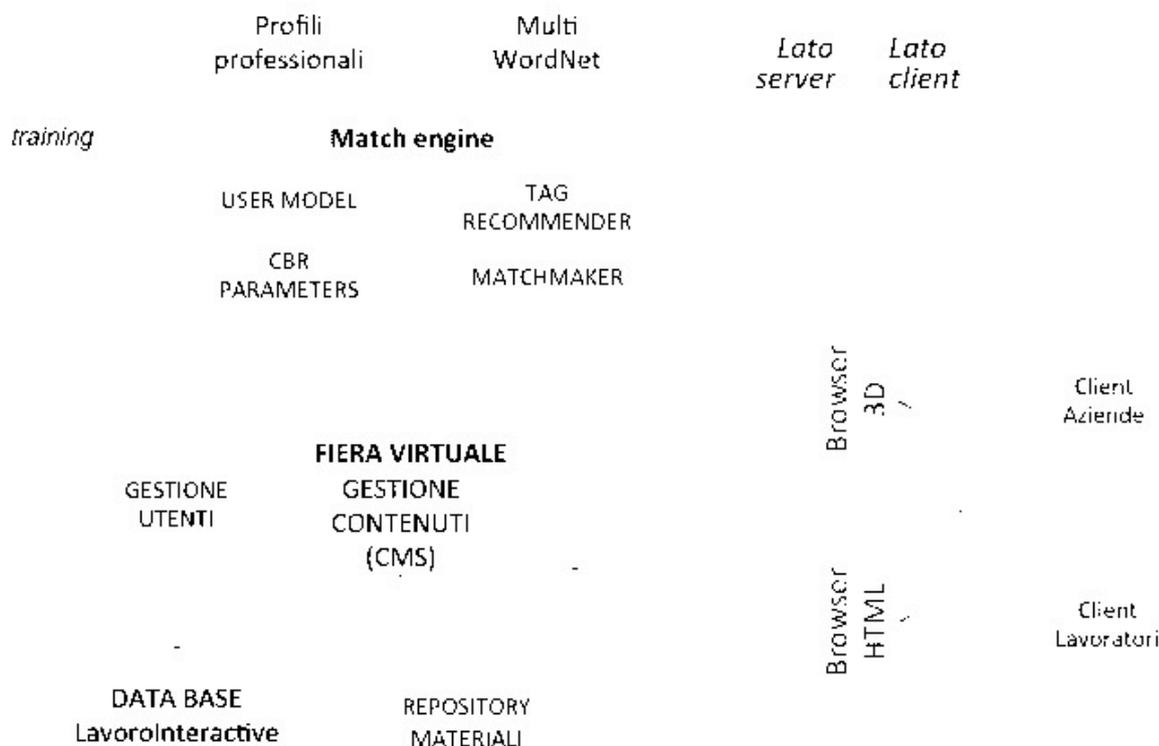


Figura 6: Architettura software di Lavoro Interactive 3.0

Il Server di Matchmaking comunica direttamente con il solo Web server, il quale funge da snodo dell'architettura. La comunicazione tra Server di Matchmaking e Web server consiste dei flussi seguenti:

Da Web server a Server di Matchmaking:

- Il modello utente viene inizializzato con i dati immessi dall'utente in fase di registrazione e aggiornato sulla base delle modifiche successive, a partire dalla funzione di gestione degli utenti nel Web server.
- Il tag recommender e il matchmaker utilizzano come input, rispettivamente, l'associazione tag-utenti (e tag-annunci) e la descrizione di candidati e annunci di lavoro, memorizzati nel database di Lavoro Interactive. In pratica, quando tali moduli vengono invocati dal Web Server, nel corso dell'interazione utente-sistema, essi "leggono" i dati presenti nel database di Lavoro Interactive.

Da Server di Matchmaking a Web server:

- Il Web Server riceve l'insieme dei tag suggeriti per un dato candidato/annuncio e l'insieme degli annunci suggeriti per un dato candidato (o dei candidati suggeriti per un dato candidato). In pratica, il Web Server richiede i suggerimenti (di tag o annunci/candidati) per un determinato annuncio/candidato, ricevendo in risposta i dati calcolati dal Server di Matchmaking.

Rispetto all'interazione utente-sistema, il **suggerimento di tag** viene reso disponibile durante la fase di inserimento delle informazioni (durante la registrazione iniziale o nel caso di modifiche successive), necessaria per usufruire dei servizi personalizzati del sito. Dopo aver inserito i propri dati personali, l'utente seleziona da un elenco strutturato (classificazione ISTAT dei profili professionali) i profili professionali che ritiene descrivano al meglio le sue professionalità. Al termine di questo passaggio all'utente viene richiesto di aggiungere dei tag al proprio profilo, che vengono suggeriti dal sistema sulla base delle informazioni aggiunte dall'utente.

L'utente che ha effettuato la procedura di login sul sistema può usufruire della personalizzazione, sia nella ricerca di annunci compatibili con il proprio profilo (**matchmaking**), sia nella ricerca tramite parole chiave. Nel primo caso, il sistema cerca nell'insieme degli annunci pubblicati sul sito quelli compatibili con il profilo utente e ordina i risultati secondo il grado di compatibilità candidato/annuncio, mostrando i primi n annunci (si veda la sezione relativa alla progettazione del matchmaker per l'ottimizzazione del parametro n). La compatibilità viene mostrata all'utente per mezzo di un indicatore grafico, indicante il grado di compatibilità; per ogni annuncio compatibile (cioè che superi una soglia stabilita come minima e dipendente come vedremo dai vincoli di obbligatorietà), all'utente viene offerta la possibilità di candidarsi per la corrispondente posizione lavorativa o all'azienda di considerare una candidatura.

Nella **ricerca per parole chiave**, gli annunci trovati sulla base delle parole chiave inserite dall'utente vengono ordinati sulla base della loro eventuale compatibilità con il profilo dell'utente. Al momento di candidarsi per un'offerta di lavoro, se la compatibilità calcolata è bassa, il sistema avvertirà l'utente di questo fatto, permettendo tuttavia all'utente di avanzare la propria candidatura. In questo modo, si garantisce che l'utente abbia sempre il controllo finale sul comportamento del sistema (può sempre candidarsi) ma non viene incoraggiato a candidarsi per gli annunci per i quali la compatibilità stimata dal sistema è bassa. A ottobre 2013, l'ordinamento e gli alert non sono ancora implementati.

La rappresentazione in termini formali del dominio che costituisce l'ambito di applicazione di Lavoro Interactive, cioè la descrizione degli utenti e delle offerte di lavoro (o annunci), è di cruciale importanza per il funzionamento del suggerimento di tag per il matchmaker.

La descrizione del profilo professionale dell'utente si basa principalmente su un insieme di etichette che rappresentano le professioni riconosciute dal sistema: *addetti alle casse, elettrotecnico*, ecc. .

La valenza di queste etichette, denominate profili professionali nella terminologia di IoLavoro, è duplice: da un lato, i profili di IoLavoro sono stati posti in corrispondenza con la **classificazione delle professioni ISTAT** stabilita a livello nazionale (<http://cp2011.istat.it>), così che ad ogni profilo di IoLavoro corrisponde un profilo della

classificazione ISTAT⁴: per esempio, il profilo *addetti alle casse* IoLavoro corrisponde al profilo *5.1.2.4.0.3 cassiere di negozio* nella classificazione ISTAT. Questa corrispondenza garantisce l'interoperabilità del sistema con la classificazione delle professioni a livello nazionale.

I profili professionali sono ulteriormente strutturati secondo una tassonomia costruita ad hoc per la classificazione di candidati e annunci in IoLavoro. La **classificazione delle professioni di IoLavoro** (Ottobre 2012) consiste di 8 settori, divisi in 19 gruppi, i quali contengono i 419 profili professionali previsti dalla classificazione. Per esempio, cassiere di negozio appartiene al gruppo *addetti vendita* e al settore *gdo/commercio*.

Il raggruppamento dei profili professionali in gruppi e settori sorge dalla necessità di ovviare a un'importante limitazione della classificazione ISTAT: i livelli superiori di tale classificazione, infatti, sono basati sul grado di qualificazione nella professione, cosicché professioni simili per l'ambito in cui si applicano vengono ad essere distanti nella tassonomia. Per esempio, *chef* fa parte della categoria 3.1.5.4.1-Tecnici della preparazione alimentare, (dove il raggruppamento 3 rappresenta le *professioni tecniche*), mentre *cuoco* fa parte della categoria 5.2.2.1.0 - Cuochi in alberghi e ristoranti (dove il raggruppamento 5 rappresenta le *professioni qualificate*).

Durante lo svolgimento delle passate edizioni della fiera del lavoro IoLavoro si è rivelato in alcuni casi proficuo cercare candidati con profili professionali correlati, anche se non uguali, a quello richiesto. Per ovviare alla distanza concettuale tra professioni simili introdotta dalla classificazione ISTAT, nella tassonomia di IoLavoro esse sono state raggruppate per gruppi e settori: *cuoco* e *chef*, per esempio, vengono a trovarsi nello stesso gruppo (*personale di cucina*) e nello stesso settore (*alberghiero/ristorazione*).

Il sistema si avvale, oltre alla conoscenza relativa al dominio specifico delle professioni, anche della conoscenza *commonsense* incorporata nel sistema WordNet. WordNet è un *database* lessicale, in cui ai termini sono associate informazioni semantiche che permettono a un sistema automatico di fare semplici inferenze a partire dai termini stessi:

- Relazioni semantiche tra termini: le relazioni di *sinonimia* (stesso significato), *iperonimia* (termini più generali), *iponimia* (termini più specifici) collegano in una "rete" tutti i termini contenuti nel lessico. Ad esempio, *parrucchiere* ha come sinonimo "acconciatore", "segretario" ha come termine più generale "assistente".
- Classificazione dei termini secondo categorie concettuali: i termini sono accompagnati dall'indicazione di uno o più "dominii" concettuali a cui appartengono. I 164 domini sono a loro volta organizzati in una tassonomia su quattro livelli. Per esempio, "segretario" appartiene ai domini "Commerce" e "Person", "parrucchiere" appartiene al dominio "Body Care".

⁴ La corrispondenza tra le professioni di IoLavoro e i profili professionali riguarda il livello più basso della classificazione ISTAT, cioè le professioni vere e proprie (i livelli superiori della classificazione ISTAT raggruppano infatti le professioni per tipologie di varia natura).

Queste informazioni sono importanti per la funzionalità di matchmaking, dato che permettono sia di ragionare sulla similarità tra i tag, utilizzando la relazione di sinonimia e quella di iperonimia/iponimia, sia di attaccare ai tag conoscenza semantica generale, utilizzando la classificazione dei termini nei domini. In Lavoro Interactive si utilizza una versione italiana di WordNet, MultiWordNet, che supporta l'interfacciamento con la versione inglese ai fini dello sviluppo di sistemi multi-lingua⁵.

Analizzando i dati di IoLavoro e il problema dell'incontro domanda-offerta nell'ambito del lavoro, il modello generale descritto in precedenza è stato parametrizzato nei termini seguenti. Consideriamo un annuncio tipo:

"RESPONSABILI MINI - BABY CLUB"

RESPONSABILI ANIMAZIONE BAMBINI IN VILLAGGI TURISTICI CLUB ESSE IN ITALIA

Profilo: Responsabili Miniclub

Gruppo: responsabili

Età (min - max): 22 - 30

Titolo di studio richiesto*: maturità / diploma

Categoria protetta: No

Tag: animazione bambini, responsabile mini club, baby club

dove sono state evidenziate in grassetto le caratteristiche (*feature*) da compilare, richieste dalla modulistica di IoLavoro (edizione Ottobre 2012). Le feature con asterisco sono obbligatorie; le rimanenti preferenziali. Alcune feature sono obbligatorie per definizione (Categoria protetta, Disabilità); altre, come nel caso di Età o Titolo di Studio, possono definite obbligatorie in un singolo annuncio a discrezione dell'azienda.

Quindi, vediamo un candidato tipo:

Titolo di studio: Maturità / Diploma

Età: 25

Profili indicati: Capo Mini Club, Responsabili Miniclub, Responsabili Animazione, RECEPTIONIST, Addetti booking, Operatore Agenzie di Viaggio

Gruppi indicati: responsabili, ricevimento

Notiamo che un candidato presenta più profili (rispetto al singolo profilo dell'annuncio).

Facendo un test preliminare sull'archivio di IoLavoro (edizione Ottobre 2012), eseguendo un match sui profili, sicuramente la feature più aderente alla ricerca, si ottiene il grafico in Figura 7.

⁵ Pianta E., Bentivogli L., Girardi C., MultiWordNet: developing an aligned multilingual database, In Proceedings of the First International Conference on Global WordNet, Mysore, India, January 21-25, 2002; Bentivogli L., Forner P., Magnini B., Pianta E., Revising the wordnet domains hierarchy: semantics, coverage and balancing, In Proceedings of the Workshop on Multilingual Linguistic Resources (LREC), pp. 101--108, ACL, 2004.



rispondere ai requisiti; la soluzione è di ordinarli secondo criteri di preferenzialità e proporli all'utente in un elenco ordinato. Cominciamo con i gruppi.

Consideriamo il seguente esempio, di annuncio e candidato rispettivamente.

Profilo: Addetto Mensa
Gruppo: personale di sala / bar

Profili indicati: Commis di Cucina, cuoco di imprese per la ristorazione collettiva, Addetto Cucina, Aiuto cuoco, Cuisinier
Gruppi indicati: personale di cucina, personale di sala / bar

Come si nota, il profilo di questo annuncio non corrisponde a nessuno dei profili indicati dal candidato, ma corrisponde a uno dei due gruppi. Allargando la rosa delle candidature ai gruppi, questo candidato sarebbe preso in considerazione. Infatti, considerando l'esempio di Animatori jolly e Animatori di contatto, che appartengono allo stesso gruppo (Animatori), si avrebbero, nel caso di IoLavoro (edizione Ottobre 2012) 334 candidati, dati 177 Animatori di contatto + 157 Animatori jolly.

Nel caso dei tag, si fa ricorso ad annotazioni degli utenti. Consideriamo i due candidati:

Titolo di studio: Maturità / Diploma
Età: 25
Profili indicati: Capo Mini Club, Responsabile Miniclub, Responsabile Animazione, Receptionist, Addetto booking, Operatore Agenzie di Viaggio
Gruppi indicati: responsabili, ricevimento
Tags: capo mini club, responsabile mini club, infanzia, animazione, mini club

Titolo di studio: Maturità / Diploma, Dottorato / Specializzazione
Età: 25
Profili indicati: Commis di Cucina, Cuoco di imprese per la ristorazione collettiva, Addetto Cucina, Aiuto cuoco, Cuisinier, Addetto Mensa
Gruppi indicati: personale di cucina, personale di sala / bar
Tags: cuoco, aiuto cuoco, stagione, impegno, amore lavoro

Come si può notare, i tag aggiungono informazioni che non sono colte dalle componenti strutturate delle descrizioni ("amore lavoro" è di sicuro una nozione trasversale, spesso desiderabile per un datore di lavoro, certamente senza un campo dove inserirla; "infanzia" andrebbe inferita da "Mini club", ma non sarebbe immediato). Far corrispondere i tag permetterebbe di ampliare notevolmente la selezione, con il rischio tuttavia di includere elementi non significativi. Come si vede più avanti, la similarità tra i tag rischia di selezionare troppi elementi, per cui la soluzione più appropriata sembra essere quella di limitare la similarità dei tag ai profili (vedi in basso). In ogni caso, si introduce una funzione di similarità simmetrica per profili, gruppi e tag di annunci/candidati:

$$Similarity(t, c) = \frac{\sum_{i=1..n} w_i \times sim(t_i, c_i)}{\sum_{i=1..n} w_i}$$

La funzione di similarità tra un *target* e un *candidate* è la sommatoria delle similarità tra le *feature* del target (t_i) e le *feature* del candidate (c_i), pesate sul numero di *feature*. La similitudine tra le *feature* si calcola poi con la funzione

$$sim(t, c) = 1 - |f_t - f_c|$$

Cioè, la similarità è data dal complemento della differenza in valore assoluto dei valori delle *feature*. Per comprendere il significato intuitivo di questa misura si pensi ai due esempi seguenti:

$$sim(t, c) = 1 - |f_t - f_c| = 1 - |0,8 - 0,7| = 1 - 0,1 = 0,9$$

$$sim(t, c) = 1 - |f_t - f_c| = 1 - |0,1 - 0,7| = 1 - 0,6 = 0,4$$

Nel caso in cui due *feature* abbiano un valore vicino (primo caso), la similarità risulta premiata, risultando in un incremento rispetto ai due valori di partenza; nel caso in cui le due *feature* abbiano un valore molto differente, la similarità risulta essere attenuata.

Infine, la similarità tra i tag è di due tipi: lessicale e semantica. Nel caso lessicale, si fanno corrispondere i termini: ad esempio, l'annuncio con i tag

capo, contatto, mini club, torneisti, dee jay, cantanti, ballerini

sarà simile al candidato con i tag

mini club, animatore, villaggi, infanzia, contatto

anche se i profili e i gruppi non coincidevano; nel caso semantico, l'annuncio con i tag

windsurf, istruttore, animatore sportivo

sarà simile al candidato con i tag

web, cuoco, sci, maestro

dove istruttore e maestro sono semanticamente simili.

Utilizzando gruppi e tag, si possono coprire le offerte rimaste scoperte con i profili. Il grafico in Figura 8 mostra come le nuove proposte riescano a superare quelle previste con i profili.



Figura 8: Test preliminare su IoLavoro Ottobre 2012, match su gruppi e tag. Si nota la quasi totale copertura degli annunci.

La soluzione quindi ipotizzata per Lavoro Interactive distingue innanzitutto tra una fase di selezione sulla base delle feature obbligatorie, che includono i profile; tale selezione si può allargare ai gruppi e ai tag per avere più candidature; gli elementi selezionati si ordinano alla similarità tra annunci e candidati risultanti dalla selezione (in questa fase, i gruppi sono risultati di poca utilità); si ha infine un ordinamento alternativo basato sulla dissimilarità (cosiddetta soluzione “Stupiscimi”, vedi avanti). E’ importante notare che il calcolo delle funzioni descritte è asimmetrico, e va calcolato sia nel caso delle candidature da ricercare per un annuncio, sia nel caso degli annunci da ricercare per un candidato.

Prima di riprendere le funzionalità di matchmaking, occupiamoci della valorizzazione del modello con i dati provenienti dagli archivi di IoLavoro (edizione Ottobre 2012); la valorizzazione viene utilizzata per determinare i pesi della funzione di similarità di cui sopra; in seguito all’applicazione delle funzioni pesate, i pesi vengono aggiustati per aumentare l’efficacia (edizione Marzo 2013). Cominciamo dal suggerimento di tag.

Come rilevato nella sezione “Lavoro Interactive 3.0: overview”, la capacità di un sistema automatico (o di un essere umano) di ragionare sulla corrispondenza tra candidati e annunci di lavoro (*matchmaking*) dipende in maniera cruciale da due fattori principali:

- l'utilizzo di uno stesso schema descrittivo tra candidati e annunci, che pone le basi per un confronto efficace: per esempio, se nella descrizione dell'annuncio compare l'elemento “titolo di studio”, lo stesso elemento dovrà comparire nella descrizione dei candidati, in modo che si possa verificare se la corrispondenza tra titolo di studio richiesto dall'annuncio e titolo di studio posseduto dal candidato sussiste;
- la qualità della descrizione contenuta nei vari elementi dello schema, lato candidati e lato annunci, e il formato con cui essa è presentata: l'utilizzo di etichette codificate per descrivere le competenze acquisite (titoli di studio, esperienze professionali, ecc.), per esempio, garantisce che possa essere facilmente rintracciata la corrispondenza tra i requisiti posti dall'annuncio e quelli di cui i candidati sono in possesso.

Mentre l'adesione agli standard nello schema descrittivo è garantita dall'adozione degli schemi sviluppati da APL per IoLavoro, la qualità della descrizione può ricevere un beneficio significativo dall'utilizzo dei tag, secondo la prassi comune del Social Web. A partire dall'ultimo quinquennio, infatti, gli utenti del Web sono stati incoraggiati ad associare ai contenuti (propri e altrui) etichette descrittive sintetiche, note come *tag*. L'insieme dei tag associati ai contenuti di un sito (noto come *folksonomia*) costituisce uno degli *asset* principali del sito stesso, dal momento che essi forniscono un'indicizzazione “dal basso” dei contenuti, vicina alle concettualizzazioni che di essi si formano spontaneamente gli utenti.

Data l'importanza dei tag per l'indicizzazione dei contenuti, in Lavoro Interactive i tag sono stati introdotti sia nello schema descrittivo dei candidati che in quello degli annunci. I tag infatti, benché compaiano talvolta nei siti commerciali a scopo informativo, cioè come complemento alla descrizione schematica degli annunci, costituiscono una risorsa importante anche per il *matchmaking*. Per questo motivo, in Lavoro Interactive 3.0 è stato integrato uno strumento di suggerimento di tag (*tag recommender*) che assiste gli utenti nella selezione dei tag da associare ai contenuti.⁶

Il tag recommender è stato sviluppato a partire dall'analisi dei fattori seguenti:

- L'insieme dei tag che gli aspiranti lavoratori hanno utilizzato per descriversi nell'edizione di IoLavoro di ottobre 2012, nel corso della quale i tag sono stati introdotti a scopo sperimentale. Sono state analizzate:

⁶ cfr. Jaeschke R., Marinho L., Hotho A., Schmidt-Thieme L., Stumme G., Tag recommendations in folksonomies, In Knowledge Discovery in Databases: PKDD 2007, pp. 506-514, Springer, 2007; Markines B., Cattuto C., Menczer F., Benz D., Hotho A., Stumme G., Evaluating similarity measures for emergent semantics of social tagging, In Proceedings of the 18th international conference on World wide web, pp. 641-650, 2009.

- la tipologia dei tag, la loro frequenza, il numero di tag per candidato;
 - le correlazioni tra tag.
- La relazione tra candidato (e annuncio) e tag. In particolare, sono state analizzate:
 - la relazione tra descrizione del candidato e tag;
 - la relazione profili professionali e tag;

Sulla base degli elementi emersi da questa analisi, descritta nelle sezioni successive (“I tag nel dataset di IoLavoro” e “Principi del suggerimento di tag”), è stato individuato come approccio idoneo al suggerimento di tag in Lavoro Interactive 3.0 la tecnica che consiste nel suggerire a ogni utente i tag utilizzati per descriversi dagli utenti a lui (o lei) maggiormente simili (cf. sezione “Il suggerimento di tag in Lavoro Interactive 3.0”). L’individuazione della similarità tra gli utenti è stata affidata a una tecnica di tipo statistico, attestata nello stato dell’arte e adattata alla descrizione dei candidati (e degli annunci) propria di Lavoro Interactive 3.0.

Un altro approccio molto comune consiste nel suggerire tag a partire da uno o più tag già associati dall’utente a una certa risorsa: tale strategia è stata ritenuta meno affidabile a seguito di una simulazione condotta sull’insieme dei tag di IoLavoro di ottobre 2012. Inoltre, essa presenta l’inconveniente di essere praticabile solo quando l’utente ha già associato a una certa risorsa almeno un singolo tag – sulla base del quale possono essere suggeriti altri tag che hanno con esso una forte correlazione.

La progettazione del sistema di suggerimento di tag per Lavoro Interactive 3.0 è stata condotta sulla base dell’analisi del tagset dell’edizione di IoLavoro di ottobre 2012.

I tag raccolti per IoLavoro (ottobre 2012) sono 9040, prodotti da 2552 utenti diversi, quindi il numero medio di tag per utente è 3,54. I tag non ripetuti sono 2838, quindi il rapporto type/token è di 3,19 occorrenze per ogni tag.

L’analisi tipologica si è basata sulle seguenti tipologie di tag:

- Oggetti concreti
- Parole in inglese
- Nomi di professione
- Nomi composti
- Parole inesistenti
- Luoghi di lavoro
- Errori ortografici
- Qualifiche professionali

Per ogni tipologia, sono state conteggiate le occorrenze di tag che appartengono alla tipologia data. La tabella seguente (Figura 9) riporta i conteggi effettuati (sul totale di 2383 tag). La presenza di una percentuale elevata di nomi composti e di nomi di

professioni hanno influenzato la progettazione delle funzionalità di suggerimento di tag e di *matchmaking*.

SETTORE	NUMERO OCCORRENZE	PERCENTUALE %
Nomi composti (esclusi nomi professioni)	1360	57%
Termini astratti	330	13,8%
Nomi di professione	310	13%
Parole in inglese	239	10%
Settore	168	7%
Errori ortografici	160	6,7%
Luoghi di lavoro	112	4,7%
Oggetti concreti	79	3,3%
Parole inesistenti	54	2,3%

Figura 9: Tipologie di tag nel tagset di IoLavoro (ottobre 2012). Alcuni tag appartengono a più tipologie. Si osservi l'alta percentuale di nomi composti.

I tag sono stati analizzati anche rispetto alla loro appartenenza ai tipi di lessemi, utilizzando le tipologie lessicali incorporate in WordNet. Questa analisi ha comportato la traduzione, condotta con metodi semiautomatici, dei tag dall'italiano all'inglese (dato che l'indice è presente solo in WordNet e non in MultiWordNet). La disambiguazione dei tag tradotti in inglese è stata fatta manualmente. La tabella in Figura 10 seguente riporta i risultati:

TIPO	# tag	percentuale
noun.person	216	26.44%
noun.act	141	17.26%
noun.artifact	116	14.20%
noun.communication	80	9.79%
noun.cognition	58	7.10%
noun.group	37	4.53%
adj.pert	28	3.43%
adj.all	25	3.06%
noun.attribute	21	2.57%
noun.location	15	1.84%
noun.state	15	1.84%
noun.process	10	1.22%
noun.food	8	0.98%
noun.time	8	0.98%
noun.event	7	0.86%
noun.phenomenon	5	0.61%
noun.possession	5	0.61%
adv.all	3	0.37%
noun.feeling	3	0.37%
noun.relation	3	0.37%
noun.substance	3	0.37%
noun.Tops	3	0.37%
adj.ppl	2	0.24%
noun.motive	2	0.24%
noun.object	2	0.24%
noun.body	1	0.12%
noun.plant	0	0.00%
noun.quantity	0	0.00%
	817	100.00%

Figura 10: Tipologie lessicali (nomi, aggettivi, verbi, ecc.) rappresentati nel tagset di Iolavoro (ottobre 2012).

Si può osservare come la prevalenza del tipo noun.person (26%) corrisponda probabilmente al numero elevato di tag che indicano professioni nel tagset (13%, il secondo gruppo più significativo dopo quello rappresentato dai termini composti, che include anche nomi di professione, per es. “cassiera”), mentre noun.act (17,26%) può essere associato ai tag costituiti dai termini astratti, tra cui ricorrono frequentemente termini indicanti l’azione tipica di una professione (“assistenza”). Similmente a quanto avvenuto per l’analisi tipologica, anche questi dati hanno influenzato il trattamento dei tag in Lavoro Interactive 3.0, orientando la progettazione verso i nomi di professione e i nomi composti in particolare (modello parametrico).

Tipicamente i sistemi sociali che usano tag per descrivere le risorse utilizzano una relazione ternaria di questo tipo:

$$\{(u, r, (t_1, \dots, t_n))\}$$

dove u indica un utente che etichetta (*tagga*) la risorsa r con i tag t_1, \dots, t_n .

Nel caso specifico del progetto Lavoro Interactive, la risorsa r che l'utente di tipo candidato tagga è in realtà l'utente stesso. Per gli utenti di tipo "azienda", l'oggetto taggato è il singolo annuncio.

Dall'analisi effettuata sul dataset ottenuto dall'edizione della fiera IoLavoro di ottobre 2012 è emersa una situazione simile a quella rappresentata nella Figura 11. Spesso l'utente usa dei tag collegati e in alcuni casi simili ai profili professionali indicati, come si evince anche dai dati presentati nella sezione precedente.

Secondo lo schema di descrizione di candidati proprio di IoLavoro (e acquisito da Lavoro Interactive), l'utente descrive se stesso indicando una serie di profili (p_1, \dots, p_n), forniti dal sistema, per i quali si candida per la ricerca di lavoro:

$$\{(u, r, (p_1, \dots, p_n))\}$$

Dunque potremmo dire che l'utente descrive se stesso usando

- una serie di profili (scelti nell'insieme di quelli disponibili, come descritto nella sezione "Basi di conoscenza in Lavoro Interactive 3.0")
- una serie di tag liberi.

L'azienda, invece, descrive un singolo annuncio, al quale corrisponde un solo profilo: il caso delle aziende, quindi, costituisce una versione semplificata del caso degli utenti di tipo candidato, perciò l'analisi condotta si concentra sul suggerimento di tag ai candidati, da cui si deriverà un modello semplificato per il suggerimento di tag alle aziende.



Figura 11. Relazione tra utente, profili e tag

Come si vede nello specifico in Figura 11, i tag T1 e T2 si riferiscono al profilo lavorativo P1, il tag T3 si riferisce al profilo P3, mentre il tag T4 non si riferisce a nessun profilo.

Per esemplificare questa relazione su casi reali, si considerino gli utenti seguenti:

- profili: "Architetto paesaggista"
- tag: "architetto", "paesaggista", "animatore"

i primi due tag sono direttamente riferiti all'unico profilo "Architetto paesaggista", mentre il terzo indica un'esperienza lavorativa compiuta che evidentemente l'utente non ha voluto utilizzare per qualificarsi professionalmente in modo esplicito attraverso i profili.

Un altro caso evidenzia come alcuni profili non vengano coinvolti nella relazione con i tag:

- profili: "Addetto al bazar/brico", "Addetto alla logistica", "Addetto alle vendite", "Magazziniere", "Scaffalista", "Vetrinista"
- tag: "magazziniere", "vendita", "spedizioni", "gestione"

In questo caso, il profilo Vetrinista non è collegato direttamente ai tag (se non, indirettamente, attraverso il termine generico "vendita"); i tag sono collegati invece al profilo "Addetto alla logistica" ("spedizioni"), "Addetto alle vendite" ("vendita"), "Magazzinieri" ("magazziniere"). "Gestione" sembra piuttosto designare il tipo di incarico ricoperto negli ambiti professionali indicati.

In sintesi, a partire dall'insieme dei profili inseriti dall'utente (p_1, \dots, p_n) si possono apprendere le associazioni che questi profili hanno con i tag che li accompagnano, e utilizzare queste associazioni apprese per suggerire dei tag a partire da profili noti. Tali associazioni sono evidenti ad un analista umano, ma un modulo software, non opportunamente istruito, non è in grado di inferire le corrette relazioni tra tag e profili, e tra i tag stessi che descrivono gli utenti.

Al fine di stabilire un'associazione tra tag e utenti, utilizzabile al fine del suggerimento, abbiamo deciso di analizzare il dataset (e in particolare le relazioni sopra descritte) usando una nota metrica di similarità, l'indice di Jaccard. L'**indice di Jaccard**, noto anche come "coefficiente di similarità di Jaccard", è un indice statistico utilizzato per confrontare la **similarità** e la **diversità** di insiemi campionari. Il coefficiente di Jaccard è definito come la dimensione dell'intersezione divisa per la dimensione dell'unione degli insiemi campionari.

$$J(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|}$$

Il coefficiente di Jaccard è stato utilizzato in diversi studi per raccomandare tag a partire da un tag conosciuto⁷. Nell'esempio sopra, la co-occorrenza normalizzata dei due tag t_i e t_j ci dirà quanto i due sono correlati in base all'utilizzo che ne viene fatto dagli utenti, ossia in base al numero di volte in cui i due tag sono usati insieme rispetto a quelle in cui sono utilizzati disgiunti. t_j

Abbiamo deciso di calcolare il coefficiente di Jaccard sui seguenti insiemi:

- A) per ogni **coppia di tag**. Il coefficiente in questo caso considera il numero di intersezioni tra i due tag all'interno del tagset di ogni singolo utente (quante volte all'interno di ogni tagset utente t_i e t_j sono usati insieme), diviso per l'unione delle occorrenze dei due tag considerati singolarmente (quante volte è stato utilizzato il tag t_i più quante volte è stato utilizzato il tag t_j)
- B) per ogni coppia **profilo/tag**. Il coefficiente in questo caso considera il numero di intersezioni tra un singolo tag e un singolo profilo all'interno del tagset dei profili (quante volte all'interno del profilo compare quel tag), diviso per l'unione delle occorrenze dei due elementi considerati singolarmente (quante volte è stato indicato quel profilo più quante volte è stato utilizzato quel tag). La relazione tra questi due insiemi è emersa come particolarmente rilevante dall'analisi dei tag, dalla quale i tag indicanti le professioni degli utenti sono risultati maggioritari.
- C) tra **ogni set di profili che descrivono un utente e tutti gli altri profili** degli altri utenti del dataset per valutare la similarità tra i profili. In questo caso specifico è stata usata una variante dell'indice di Jaccard per la similarità tra vettori, chiamata **distanza di Tanimoto**. In questo caso, l'utente è visto come descritto da un vettore di profili, $u = [p_1 \dots p_n]$ e dunque viene calcolata la similarità con tutti gli altri vettori di profili che descrivono gli utenti utilizzando il rapporto tra l'intersezione dei vettori diviso la loro unione.

Il primo caso (indice di similarità tra i tag, punto a) è stato analizzato per fare emergere le relazioni di similarità tra i tag presenti nel tagset. Il coefficiente di Jaccard permette, in questo caso, di scoprire le relazioni di similarità tra i tag così come emergono dal comportamento degli utenti. Più l'indice di Jaccard è alto, maggiore sarà la similarità. L'ipotesi di lavoro sottostante a questa correlazione è che, utilizzando l'indice di Jaccard tra coppie di tag è possibile individuare i tag da raccomandare a partire da un tag inserito dall'utente, secondo un semplice algoritmo:

⁷ Sigurbjörnsson B., van Zwol R., Flickr tag recommendation based on collective knowledge. In Proceedings of the 17th international conference on World Wide Web (WWW '08). ACM, New York, NY, USA, 327-336, 2008. DOI=10.1145/1367497.1367542 <http://doi.acm.org/10.1145/1367497.1367542>.
Cattuto C., Benz D., Hotho A., Stumme G.: Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. International Semantic Web Conference, pp. 615-631 2008.



dato l'insieme dei tag utente (t_1, \dots, t_n)

1. per ogni tag utente t_i
 - a. per ogni tag t_j nel dataset
 - i. calcola l'indice di Jaccard tra p_i e t_j
2. suggerisci gli n tag con indice di Jaccard più alto

Il calcolo di Jaccard sulle possibili coppie di tag può essere effettuato offline e memorizzato in una tabella che il tag recommender consulterà quando necessario.

Es. dato il tag *barista*, i tag con indice di Jaccard più alto, e dunque i primi 5 tag maggiormente simili in base all'indice di Jaccard sono:

Tag	Indice di Jaccard
Cameriere	0.14680850505828857
Commessa	0.07189542800188065
Aiuto cuoco	0.0385604128241539
Scaffalista	0.036827195435762405
Animatrice	0.03478260710835457

Dato il tag *segretaria*, i tag con indice di Jaccard più alto sono:

Tag	Indice di Jaccard
Commessa	0.10559006035327911
Impiegata amministrativa	0.05590062215924263
Receptionist	0.0555555559694767
Cassiera	0.05106382817029953
Barista	0.04337349534034729

Dato il tag *magazziniere*, i tag con indice di Jaccard più alto sono:

Tag	Indice di Jaccard
Scaffalista	0.1071428582072258
Operaio	0.1042780727148056
Cameriere	0.07064017653465271
Carrellista	0.06643356382846832
Autista	0.06624605506658554

Come si può notare dagli esempi, tuttavia, i tag simili secondo l'indice di Jaccard, non sempre risultano essere semanticamente simili. L'analisi dei suggerimenti ottenuti con l'algoritmo sopra descritto ha messo in luce che, nel caso di IoLavoro, l'indice di Jaccard riflette il reale comportamento di *tagging* degli utenti, che spesso sono caratterizzati da esperienze lavorative molto eterogenee. Quindi ad esempio il fatto che il tag *magazziniere* sia spesso associato a *cameriere* indica che gli utenti che hanno svolto entrambe le professioni sono molti.

Dunque suggerire i tag solo in base alla loro co-occorrenza non è una strategia vincente, perché la scarsa similarità semantica tra i tag inseriti dall'utente e i suggerimenti



ottenuti potrebbe confondere l'utente. Abbiamo quindi ipotizzato che una maggiore significatività potesse emergere dalle relazioni tra tag e profili.

Dal momento che i profili sono usati dall'utente per descrivere se stesso, così come lo sono i tag, abbiamo analizzato la co-occorrenza tra ogni singolo profilo e i tag, associazione che dovrebbe maggiormente riflettere una similarità semantica. L'indice di Jaccard calcolato tra profili e tags (punto b) è stato quindi utilizzato per evidenziare i legami tra tag e profili singoli, così come emergono dal dataset. Più l'indice di Jaccard è alto tra un dato tag e un dato profilo, più quel tag sarà vicino a quel profilo.

Il suggerimento, quindi, potrebbe essere gestito dall'algoritmo seguente:

dato l'insieme di profili (p_1, \dots, p_n) ,

1. per ogni profilo p_i
 - a. per ogni tag t_j nel dataset
 - i. calcola l'indice di Jaccard tra p_i e t_j
2. suggerisci gli n tag per i quali l'indice è più alto.

Ad esempio, il tag `animatore` è legato a questi profili, di cui 4 sono fortemente legati al significato del tag:

Profilo	Indice di Jaccard
Animatore	0.065445026178
Animatore di contatto	0.0514905149051
Organizzatore tornei	0.0460526315789
Animatore Junior Club	0.0437788018433
Animatore Miniclub	0.0368932038835

Il tag `barista` mostra un indice di Jaccard elevato per i seguenti profili:

Profilo	Indice di Jaccard
Barista	0.119214586255
Barman	0.0838095238095
Addetto bar	0.0764748725419
Cameriere	0.0665766981556
Addetto Sala	0.0613682092555

Al contrario, dato un profilo, l'analisi dei tag con indice di Jaccard più elevato rivela i tag che hanno una maggiore associazione con quel profilo. Per esempio, dato il profilo `Barista`, i tag con indice di Jaccard più alto sono:

Tag	Indice di Jaccard
Barista	0.119214586255
Cameriere	0.0557515878617
Cameriera	0.0439798125451
Magazziniere	0.0213827512473
Commessa	0.0207587687903

Sul profilo *Segretaria*, con indice di Jaccard più alto sono i tag:

Tag	Indice di Jaccard
Segretaria	0.0390920554855
Commessa	0.023720349563
Impiegata	0.0209580838323
Cameriera	0.0162975344756
Cassiera	0.0147866497676

Sul profilo *Magazziniere*, i tag con indice di Jaccard più alto sono:

Tag	Indice di Jaccard
Magazziniere	0.100689655172
Cameriere	0.0328917313842
Operaio	0.032080659945
Barista	0.0272975432211
Scaffalista	0.0203309692671

Questa strategia, tuttavia, risente, come la precedente, del fatto che gli utenti assommano nei profili (e quindi taggano) esperienze lavorative diverse. Come si può notare dagli esempi sopra riportati, spesso i singoli profili sono associati a dei tag che non hanno nessun legame semantico con quel profilo. Per esempio, nel caso del profilo *Magazziniere*, il secondo tag per valore dell'indice di Jaccard è *Cameriere*. Questo è dovuto al fatto che gli utenti non usano solo un profilo per descriversi, ma usano un insieme di profili, così come usano un insieme di tag. Quindi, se la maggior parte degli utenti che ha nel proprio modello utente il profilo p' ha anche il profilo p'' e ha taggato p' con il tag t' e p'' con il tag t'' , la correlazione calcolata tra p' e t' sarà altrettanto forte che quella tra p' e t'' (e viceversa per p'').

Quindi, quello che emerge dalla valutazione delle strategie ipotizzate è che non un singolo profilo, ma un *set di profili* risulta avere una relazione con un *set di tag*. Questa relazione può essere rappresentata dalla seguente matrice (Figura 12), che descrive un certo utente $u1$:

u1	s3		s4
	g1		g2
	p1	p2	p3
t1	X	X	X
t2	X	X	X
t3	X	X	X
t4	X	X	X

Figura 12: Rappresentazione di un singolo utente come relazione tra tag (colonna a sx) e profili (terza riga), raggruppati in gruppi (seconda riga) e settori (prima riga).

Ogni utente del sistema si descrive con un set di tag (es. t_1, t_2, t_3, t_4) e con un set di profili (p_1, p_2, p_3), raggruppati in due gruppi e poi in due settori. Implicitamente ogni tag usato dall'utente u_1 è legato ad ognuno dei profili (p_1, p_2, p_3) ed ogni profilo risulta avere un legame con quei tag. Tale legame però ha senso solo considerando l'insieme dei profili e non un singolo profilo. Ad esempio il profilo *Barista* può avere un legame con i tag *bar* e *chef*, perché l'utente in questione ha indicato oltre al profilo *Barista* anche il profilo *cuoco* e poi ha usato quei tag per descriversi. Dunque solo considerando le relazioni tra set di profili e set di tag che gli utenti hanno indicato questi legami possono diventare significativi.

Si noti inoltre che i profili possono essere raggruppati in gruppi e settori (come illustra la figura), dunque si può sfruttare anche questo legame, come vedremo in seguito, per studiare queste associazioni.

La strategia basata su questo modello si focalizza sull'indice di Jaccard tra set di profili e set di tags (rispettivamente, le righe e le colonne della matrice). Quindi, si è pensato di calcolare la similarità tra i vettori dei profili che descrivono gli utenti (terzo punto, c) e in base a tale similarità suggerire i tag utilizzati da utenti simili ad un dato utente.

L'algoritmo di raccomandazione è il seguente:

Dato un utente x con un insieme di profili $P_x = \{p_{x,1}, \dots, p_{x,n}\}$:

1. Trova l'insieme U degli utenti y_i che hanno almeno un profilo in comune con x , ($P_x \cap P_{y_i} > 1$)
2. Per ogni utente y_i in U :
 - i. Calcola similarità con utente x (metrica di Tanimoto tra $\{p_{x,1}, \dots, p_{x,n}\}$ e $\{p_{y_i,1}, \dots, p_{y_i,n}\}$)
3. Ordina gli utenti y_i in U per valore di similarità tw , $U^{ranked} = \{<y_1, tw_1>, \dots, <y_n, tw_n>\}$
4. Prendi i primi m utenti in U^{ranked}
5. Prendi i tag dei primi n utenti in U^{ranked} , $T = \langle t_{y_1}^1, \dots, t_{y_m}^1, \dots, t_{y_1}^n, \dots, t_{y_m}^n \rangle$
6. Ordina i tag in T per numero di occorrenze $T^{ranked} = \{<t_1, f_1>, \dots, <t_n, f_n>\}$
7. Suggestisci a x i primi m tag in T^{ranked}

Nell'attuale implementazione, n (numero di utenti considerati per similarità) è fissato a 10, m (numero di tag suggeriti) è fissato a 5. Si noti, inoltre, che il passo 1 dell'algoritmo viene effettuato a scopo di ottimizzazione, dato che consente di escludere dall'insieme degli utenti considerati per la similarità quelli che non hanno alcun profilo in comune con l'utente dato.

Di seguito una tabella (Figura 13) con alcuni esempi di suggerimento di tag (colonna 3, "Tag suggeriti") a un utente caratterizzato da un determinato insieme di profili (colonna 1, "Profili indicati da utente X") in base all'algoritmo sopra descritto. Si noti che la tabella riporta anche i tag effettivamente utilizzati da ogni utente (colonna 2, "Tag dell'utente X") che permette di verificare se la strategia di suggerimento ha prodotto tag uguali o simili a quelli che l'utente ha inserito nel dataset originale. Gli esempi sono ordinati in maniera decrescente in base alla percentuale di copertura (fq.3, fq. 2, ecc.)

Profili indicati da utente X	Tag dell'utente X	Tag suggeriti
"hostess congressuale", "hostess di terra", "Camerieri"	hostess, cameriera	1) cameriera (fq. 3) 2) hostess (fq. 2) 3) promoter (fq. 2) 4) greenwork (fq. 1) 5) educatrice (fq. 1)
"Addetti alle casse", "ADDETTI ipermercato", "addetti al bazar/brico", "addetti al reparto tessile", "ADDETTI reparto elettrodomestici", "Addetto alla logistica", "Addetto alle vendite", "Cassieri", "Magazzinieri", "Scaffalista"	magazziniere, cassiere, commesso	1) magazziniere (fq. 4) 2) cassiere (fq. 2) 3) auto (fq. 1) 4) collaudatore (fq. 1) 5) fattorino (fq. 1)
"Accompagnatori turistici", "Addetti Front Office", "Agenti di viaggio", "Guide turistiche", "Organizzatori eventi", "Addetti booking", "Consulenti viaggi", "Event e Tour planner", "hostess congressuale", "tecnico di agenzia di viaggio"	agente di viaggio, guida turistica, accompagnatrice turistica, animatrice, receptionist	1) guida turistica (fq. 2) 2) receptionist (fq. 2) 3) banconista (fq. 1) 4) degustazioni vini (fq. 1) 5) accoglienza (fq. 1)
"BARMAN", "Bartender", "Capo Barman", "Bar manager", "Food & Beverage Manager"	capobarman, bar-manager, barman, bartender	1) cameriere (fq. 5) 2) barman (fq. 4) 3) operaio (fq. 2) 4) capobarman (fq. 1) 5) barista. (fq. 1)
"Operatori call center", "Operatori call center - IOLAVOROH", "hostess congressuale"	baby sitter, commessa, infanzia, animazione	1) commessa (fq. 2) 2) infanzia (fq. 1) 3) segretaria (fq. 1) 4) impiegata (fq. 1) 5) accoglienza (fq. 1)
"Segretari di Ricevimento", "RECEPTIONIST", "tuttofare d'albergo", "Addetti Front Office", "Hostess / Steward per navi", "hostess congressuale", "tecnico di agenzia di viaggio"	ricevimento, segretaria	1) receptionist (fq. 4) 2) affidataria (fq. 1) 3) educatrice prima infanzia (fq. 1) 4) segretaria (fq. 1) 5) stage in agenzia di viaggi (fq. 1)

Figura 13: Esempi di suggerimento di tag secondo l'algoritmo di tag recommendation di Lavoro Interactive 3.0.

Per valutare sistematicamente la validità della strategia proposta, è stata calcolata la sovrapposizione tra i tag che sarebbero stati suggeriti e quelli effettivamente inseriti dagli utenti (intersezione tra i due insiemi), utilizzando il tagset di IoLavoro (ottobre 2012). Per ogni tag corrispondente, il suo punteggio è stato moltiplicato per il numero di occorrenze di quel tag nel set dei suggerimenti, al fine di premiare la coincidenza tra un tag effettivamente inserito dall'utente e la "forza" con cui quel tag è suggerito.

La metrica di valutazione descritta è stata utilizzata per valutare i risultati dell'algoritmo di *recommendation* applicato all'intero dataset. Gli utenti che hanno inserito un tag sono 2554, quindi la valutazione riguarda 2554 casi. Il suggerimento presenta una coincidenza, anche parziale, con i tag presenti nel dataset in 645 casi su 2554 (25,25%)⁸. Il punteggio medio dei suggerimenti ottenuti, calcolato con la metrica sopracitata, è uguale a 9,0873. Nel complesso, questa performance è stata giudicata promettente come prima approssimazione, considerando anche il fatto che il numero di tag per utente, nel dataset, è solo di 3,54. Sulla base della valutazione effettuata, questa strategia è stata scelta come *baseline*.

Sulla base di tale strategia, è stata sviluppata una strategia semplificata per il suggerimento di tag agli utenti di tipo azienda, relativamente agli annunci. Nel caso degli annunci, non è necessario calcolare la distanza di Tanimoto tra la risorsa da taggare (l'annuncio) e le altre risorse dello stesso tipo (gli altri annunci): dato che l'insieme di profili associati alla risorsa da taggare consiste in un solo profilo, tutti gli annunci con lo stesso profilo di quello per cui si calcolano i suggerimenti avranno stesso valore di similarità per l'annuncio dato. All'utente di tipo azienda saranno quindi suggeriti tutti i tag associati agli annunci con profilo uguale, ordinati per frequenza decrescente.

Inoltre, a partire dal tagset raccolto nell'edizione di IoLavoro di marzo 2013, nel corso del quale è stato utilizzato il *tag recommender*, è stata condotta un'analisi quantitativa per verificare gli effetti del suggerimento di tag. L'analisi condotta sul tagset di IoLavoro di marzo 2013 rivela un incremento significativo nell'uso dei tag. Nell'edizione di IoLavoro di marzo 2013 i tag complessivi sono 19326 (5482 quelli non ripetuti), con una media di 5,47 tag per utente, rispetto ai 3,54 dell'edizione di ottobre 2012). La deviazione standard nel numero di tag per utente (2,27) rivela che l'aumento nell'uso dei tag ha influenzato tutti gli utenti, con lievi discrepanze tra gli utenti.

Alcuni tag, inoltre, si ripetono molto più frequentemente di altri, come si evince dalla tabella sottostante: il 19% dei tag, per esempio, occorre più di 100 volte, il 35% occorre un numero di volte superiore a 10. Tuttavia, il gruppo più numeroso è dato dai tag che occorrono meno di 10 volte (44,93%), a significare che, con l'aumento del tagset, gli utenti continuano a introdurre nuovi tag (Figura 14).

⁸ Il dato ottenuto costituisce una stima per difetto della recall dei tag presenti nel dataset da parte della strategia considerate, poiché non tiene conto del fatto che molti tag molto frequenti, diversi nella forma maschile e femminile (cameriere, cameriera), non sono considerati uguali nel match.

Numero di occorrenze del tag	tag	Numero occorrenze nel gruppo	% totale tag
100 < 1000	21	3822	19,77 %
10 < 100	234	6819	35,28 %
1 < 10	5218	8685	44,93 %
1 occorrenza	3867	3867	20,01 %

Figura 14: Raggruppamento dei tag per classi di frequenza. Il gruppo più numeroso è quello dei tag che occorrono tra 1 e 10 volte nel tagset.

Dato che la strategia di tag adottata necessita di un set di tag come base, il suggerimento di tag lato aziende (per gli annunci) non è stato applicato nell'edizione di IoLavoro di marzo 2013. Per questo motivo, non è disponibile allo stato attuale una valutazione ex-post di questo tipo di suggerimento.

La strategia scelta come baseline può essere migliorata, utilizzando ulteriori sorgenti di conoscenza o tecniche:

- Altre tecniche statistiche;
- Informazioni che possono essere inferite dalla rappresentazione dei profili professionali nella classificazione di IoLavoro;
- Informazioni che derivano dall'interazione tra utenti candidato e utenti aziende tramite il sistema;
- Conoscenza semantico-lessicale incorporata in MultiWordNet.

Mentre le prime due sono state implementate e verificate sul testset (con differenti risultati), le ultime due sono state progettate ma non implementate.

Si è pensato subito di raffinare le tecniche statistiche usate, imponendo una direzione alla similarità. L'idea è che la asimmetria può dare informazioni più accurate della simmetria, che porta a confondere chi influenza cosa. E' stata calcolata una misura asimmetrica di co-occorrenza normalizzata secondo la frequenza di uno dei tag:

$$P(t_i | t_j) := \frac{|t_i \cap t_j|}{|t_i|}$$

Tuttavia, questo coefficiente ha dato risultati meno efficaci del coefficiente di Jaccard. Sono state fatte inoltre alcune prove con algoritmi di collaborative filtering - per il

suggerimento dei tag a partire dai singoli profili - basati sulla fattorizzazione di matrici⁹, ma anche questo non ha dato risultati interessanti.

Le informazioni provenienti dalla tassonomia dei profili professionali di IoLavoro (rappresentata nel modello dell'utente dalla distribuzione dei profili indicati dall'utente nei settori e nei gruppi della tassonomia) ha migliorato la strategia di raccomandazione dei tag. Infatti, la distribuzione dei profili dichiarati dall'utente nei gruppi e nei settori permette di inferire la maggiore focalizzazione dell'utente per un certo gruppo o settore (vedi più avanti il matchmaking). Questo dato, rappresentato nel modello utente, può essere usato per integrare la misura di similarità tra gli utenti calcolata sugli insieme di profili che li descrivono.

La variante proposta procede identificando primariamente gli utenti che hanno la stessa focalizzazione dell'utente in termini di settori¹⁰, quindi applica la misura di similarità di Tanimoto tra profili a questi utenti. Seguendo questo approccio, si focalizza la ricerca di utenti simili sulla similarità inferibile dai settori prevalenti, ottimizzando al tempo stesso l'algoritmo di suggerimento (dato che Tanimoto verrà calcolato non più su tutti gli utenti che hanno almeno un profilo in comune con l'utente dato, ma solo su quelli che hanno una distribuzione simile dei profili nei settori). Per semplicità, la versione proposta prende in considerazione gli n settori che, nel modello utente di ogni candidato, hanno punteggio più alto (settori prevalenti). L'algoritmo sarà quindi il seguente:

Dato un utente x , con un insieme di profili $P_x = \{p_{x,1}, \dots, p_{x,n}\}$

1. Calcola distribuzione dei profili P_x nei settori S assegnando un peso w ai profili, $S_x = \{<s_1, w_1>, \dots, <s_n, w_n>\}$
2. Determina gli n settori prevalenti data S_x , $prevalenti(S_x) = S_x^{best}$
3. Trova l'insieme U degli utenti y_i che hanno almeno un profilo in comune con x , ($P_x \cap P_{y_i} > 1$)
4. Trova in U gli utenti che hanno stessi settori prevalenti, i.e., U' degli y_i' tali che $prevalenti(P_{y_i'}) = S_i^{best}$ dove $S_{y_i'}^{best} = S_x^{best}$
5. Per ogni utente y_i' in U' :
 - i. Calcola similarità con utente x (metrica di Tanimoto tra $\{p_{x,1}, \dots, p_{x,n}\}$ e $\{p_{y_i',1}, \dots, p_{y_i',n}\}$)
6. Ordina gli utenti y_i' in U' per valore di similarità tw , $U'^{ranked} = \{<y_1', tw_1>, \dots, <y_n', tw_n>\}$
7. Prendi i primi m utenti in U'^{ranked}
8. Prendi i tag dei primi n utenti in U'^{ranked} , $T = \langle t^{y_1'}_1, \dots, t^{y_1'}_m, \dots, t^{y_n'}_1, \dots, t^{y_n'}_m \rangle$
9. Ordina i tag in T per numero di occorrenze $T^{ranked} = \{<t_1, f_1>, \dots, <t_n, f_n>\}$
10. Suggestisci a x i primi m tag in T^{ranked}

⁹ Koren Y., Bell R. M., Volinsky C., Matrix Factorization. Techniques for Recommender Systems. *IEEE Computer* 42(8), pp. 30-37, 2009.

¹⁰ Per questa variante si considerano solo i settori (livello I della tassonomia) perché alcuni profili non sono contenuti in nessun gruppo (II livello della tassonomia).

Ogni profilo riceve un peso che dipende dall'appartenenza a un settore (settori con più profili ricevono un peso superiore) e si considerano gli n settori prevalenti; si recuperano gli utenti con almeno un profilo nell'insieme di profili dato (cioè hanno almeno un profilo in comune con l'utente) e gli utenti con gli stessi settori prevalenti. Si calcolano gli utenti simili e si prendono i primi m . Quindi, una volta presi i tag dei primi n utenti, si ordinano per numero di occorrenze e si propongono all'utente i primi m tag.

Per quanto riguarda i miglioramenti che derivano dal comportamento degli utenti, si è deciso di promuovere, nella rosa degli utenti simili, quelli che hanno avuto feedback positivi dalle aziende. Dato che la finalità del *tagging* è di migliorare l'incontro tra domanda e offerta di lavoro, l'efficacia nella scelta dei tag può essere misurata attraverso la manifestazione di interesse da parte delle aziende. Nel nostro caso, la misura di successo dei tag viene considerato il fatto che il profilo del candidato sia stato visualizzato dalle aziende, o che il candidato sia stato selezionato per un colloquio.

Questi dati, memorizzati nel modello utente, potrebbero essere a disposizione del recommender; i tag di tali utenti sarebbero preferiti rispetto agli altri, in modo che, attraverso l'interazione tra aziende e candidati, il recommender apprenda le associazioni profili/tag di maggiore successo. All'algoritmo precedentemente illustrato, quindi, si aggiungerebbe un passo in cui alla lista degli utenti ordinati per similarità con quello dato si applica un moltiplicatore sulla base del valore dell'attributo "utente selezionato" (che misura il successo dell'utente nell'ottenere contatti da parte delle aziende).

Infine, sono state effettuate analisi del dataset utilizzando MultiWordNet, per verificare la possibilità di usare questa risorsa nel miglioramento della strategia di suggerimento di tag. L'ipotesi di lavoro consiste nell'utilizzo della sinonimie codificate in MultiWordNet per espandere il significato dei tag ottenuti tramite l'approccio essenzialmente statistico sopra descritto. L'utilizzo di MultiWordNet permetterebbe inoltre di incrementare i tag presenti nel sistema (la cosiddetta "folksonomia") con nuovi tag, rendendo più efficace il suggerimento.

La limitazione principale nell'uso di questa risorsa per la recommendation è data dalla scarsa copertura dei tag contenuti nella folksonomia da parte di MultiWordNet, che si riflette anche sull'utilizzo di WordNet Domains¹¹. La percentuale di tag che corrispondono a un lemma in MultiWordNet, infatti, è limitata al 22,3% circa. Tale percentuale può essere leggermente innalzata attraverso l'utilizzo di un lemmatizzatore, senza però che questo accorgimento possa introdurre miglioramenti sostanziali nella copertura. La scarsa copertura, infatti, come evidenziato nel documento di analisi dei tag di Novembre 2012, è dovuta principalmente all'uso di parole composte (57%) e di termini stranieri (10%), piuttosto che di forme flesse.

¹¹ La gerarchia dei domini WordNet può essere consultata, nella versione incorporata in MultiWordNet, alla pagina <http://wndomains.fbk.eu/hierarchy.html>.

L'analisi delle sinonimie ricavate da MultiWordNet ha evidenziato la presenza di termini pertinenti rispetto al tag originale. Per esempio, il tag `decoratore` ha come sinonimo `tinteggiatore`, `educatore` ha come sinonimi `formatore`, `pedagogo` ha come sinonimo `insegnante`.

Un problema significativo dell'utilizzo delle relazioni sinonimiche è dato dal fatto che i tag sono polisemici: per esempio, il tag `cameriere` corrisponde a due distinti significati in MultiWordNet (tra parentesi il dominio semantico di riferimento):

1. `cameriere` (Commerce) [a person whose occupation is to serve at table (as in a restaurant)]
2. `persona_di_servizio`, `inserviente`, `servitore`, `cameriere`, `famiglio` (Person) chi svolge su incarico lavori domestici; "contributi da versare se si assume una persona di servizio"

Nel primo caso, il significato è quello che corrisponde alla figura professionale "camerieri di ristorante/bar/sala" nella classificazione di IoLavoro, mentre nel secondo caso la figura di riferimento è quella del collaboratore/collaboratrice familiare (non presente in IoLavoro). L'analisi del dataset mostra che l'ambiguità semantica costituisce la normalità nel dominio considerato: rispetto ai 507 tag presenti nel dataset e che hanno un corrispettivo in MWN, ogni tag corrisponde in media a 2,35 significati diversi.

L'analisi del livello di ambiguità semantica dei tag mostra che la strategia che consiste nell'uso dei sinonimi per il suggerimento di tag non è praticabile. Una soluzione possibile e nota in letteratura¹² consiste nell'utilizzo dell'informazione contenuta nei domini associati ai vari significati di un lemma (o tag) per capirne il significato nel contesto. Nell'esempio sopra riportato, si può notare come il dominio associato ai due significati del termine `cameriere` (`cameriere` di attività commerciale e `collaborare domestico`) siano caratterizzati da domini diversi (*Commerce* nel primo caso e *Person* nel secondo). Anche se in questo esempio l'associazione tra l'ambito concettuale di IoLavoro e i due domini considerati può sembrare intuitiva, un algoritmo che sfruttasse l'informazione contenuta nei domini per selezionare il significato corretto del tag dovrebbe essere informato da una rappresentazione formale della corrispondenza tra i profili professionali di IoLavoro e i domini semantici di MultiWordNet.

Per ottenere una rappresentazione di questo tipo, i tag presenti nel dataset sono stati analizzati sistematicamente rispetto al dominio semantico di appartenenza. L'analisi è stata compiuta su un campione molto limitato di 214 tag, dato che dei 507 tag disponibili (in quanto presenti in MultiWordNet), 293 appartengono al dominio Factotum, il quale non ha una semantica propria (e quindi trasferibile alle categorie professionali di Lavoro Interactive). In primis, i tag sono stati raggruppati per domini di appartenenza, per verificare se la loro distribuzione nei domini fosse coerente con la semantica dei domini stessi. 105 domini sono risultati rappresentati nel tagset di IoLavoro (di cui 21 con un solo tag associato e 14 con due tag associati). Con alcune eccezioni, la distribuzione dei

¹² Magnini B., Strapparava C., Pezzulo G., Gliozzo A., The Role of Domain Information in Word Sense Disambiguation, Natural Language Engineering, Special Issue on Word Sense Disambiguation, 8(4), pp. 359-373, Cambridge University Press, 2002.



tag nei domini MultiWordNet sembra corretta rispetto alla semantica dei domini stessi. Per esempio, si considerino i domini seguenti:

DOMINIO	ASSOCIAZIONE CON I TAG PRESENTI NEL DATASET
Architecture:	architetto disegnatore arredatore restauratore
Theatre:	teatro attrice spettacolo comparsa attore interprete figurante recitazione
Gastronomy:	cuoco cucina chef alimentare cuoca enologo cucinare panificatore

Un'eccezione è data dai domini più generali, quali *Person* o *Psychological_Features* o *Biology*, che, seppur rappresentati nel tagset, non hanno una semantica chiara riflessa nei tag di IoLavoro.

L'analisi della distribuzione dei tag nei domini è preliminare a ogni tentativo di associare i domini alla classificazione dei profili professionali di IoLavoro. L'attuale consistenza del set di tag collegati ai domini MultiWordNet non permette di utilizzarli per apprendere un'associazione tra domini e profili (oppure gruppi o settori), necessaria per risolvere il problema della disambiguazione del significato dei tag sopra descritto. A fronte di una ulteriore raccolta di dati, tuttavia, si potranno utilizzare tecniche statistiche per apprendere l'associazione domini-profilo (oppure gruppi o settori). In alternativa, è possibile procedere manualmente a stilare una tabella di associazione domini/profilo (o settori) che informi almeno parzialmente la strategia di disambiguazione del tag recommender. Già ora, alcuni domini specifici (quali ad esempio *Gastronomy*, *Tourism*, *Body_Care*), rivelano, attraverso i tag ad essi collegati, una semantica vicina a quella dell'ambito delle professioni di IoLavoro.

Questa sezione descrive la procedura di *matchmaking* utilizzata in Lavoro Interactive 3.0. Tale procedura permette al sito di superare i limiti propri della classica “ricerca nel database” che caratterizza la ricerca di lavoro online nei portali commerciali. I portali commerciali, infatti, mettono a disposizione degli utenti funzionalità di ricerca più o meno avanzate, senza però mettere loro a disposizione uno strumento automatico e personalizzato per l’associazione tra domanda e offerta. In particolare, l’approccio di Lavoro Interactive è centrato sull’utente, cioè l’incontro domanda-offerta è visto dalla prospettiva dell’utente, sia esso un candidato o un’azienda.

Alla base della procedura di *matchmaking* vi è il paradigma del ragionamento basato sui casi (*case based reasoning* - CBR)¹³. Il ragionamento basato su casi lavora sulla ricerca di soluzioni mediante il recupero di casi di successo da applicare a una nuova situazione ritenuta simile al caso memorizzato; nel caso del *matchmaking* tra annunci e candidati, il CBR è un tipo di *content-based filtering*, un filtro basato sui contenuti per il quale le informazioni sono rappresentate in maniera strutturata e ogni candidato è considerato come un caso da rapportare all’annuncio (e viceversa se si assume la prospettiva dell’annuncio). Il CBR è ideale per le ricerche strutturate (cioè con campi fissati) e *single-shot*, ricerche non esplorative, come nel nostro progetto.

Allo scopo di applicare il paradigma del *case based reasoning* all’incontro tra domanda e offerta di lavoro, è stata in primo luogo analizzata la corrispondenza tra lo schema descrittivo degli annunci e quello dei candidati, individuando, con la collaborazione degli esperti, i criteri di confronto maggiormente idonei tra i due schemi. Poiché il *matchmaking* si pone all’intersezione tra due funzionalità assai rilevanti nel Web attuale, cioè il suggerimento di contenuti e i motori di ricerca, al modello di corrispondenza candidato-annunci (e annuncio-candidati) ottenuto è stato associato un sistema di valutazione della rilevanza degli annunci trovati per un certo candidato (e viceversa, dei candidati trovati per un certo annuncio). In tal modo, il comportamento del *matchmaking* acquisisce un tratto tipico del motore di ricerca, contribuendo ulteriormente a orientare l’utente nella sua ricerca di lavoro (o, per le aziende, nella ricerca di candidati idonei).

Il modello teorico di match e i parametri di rilevanza sono stati valutati, in collaborazione con APL, ALL e Regione VdA, con una sperimentazione condotta sui dati (in forma anonima) raccolti durante l’edizione di IoLavoro di marzo 2013. In tale edizione, infatti, lo schema descrittivo di candidati e annunci era pienamente congruente con il modello generale, dato che in entrambi era stato introdotto l’uso dei tag. A seguito di una prima fase di valutazione, il modello è stato rivisto e i parametri corretti in modo che l’ordinamento di candidati e annunci risultante dal *matchmaking* rispecchiasse il giudizio degli esperti. A seguito di questa prima fase di ottimizzazione, una seconda fase ha permesso di convalidare il modello finale e effettuare ulteriori lievi aggiustamenti.

¹³ Smyth B., Case-Based Recommendation, The Adaptive Web, pp. 342-376, 2007.

Smyth B., Briggs P., Coyle M., O’Mahony M. P., A case-based perspective on social web search, In McGinty, L. and Wilson, D. C. (eds.). Case-Based Reasoning Research and Development : 8th International Conference on Case-Based Reasoning, ICCBR 2009, Seattle, USA, July 20-23, Springer, 2009.

Nel CBR la rappresentazione strutturata degli elementi assume la forma di liste di coppie *attributo, valore* (<*feature, value*>). In Lavoro Interactive, gli elementi sono gli annunci e i candidati. I valori predefiniti possono essere numerici o letterali. Di seguito, viene descritta in dettaglio la strutturazione di annunci e candidati per il matchmaking.

I dati riguardanti gli annunci sono stati suddivisi in *feature descrittive* (Profilo, Gruppo, Tag) e *feature preferenziali* (titolo di studio, range età, disabilità, categoria protetta). Le feature sono state derivate dall'edizione di Ottobre 2012 di IoLavoro.

I valori riguardanti tutte le *feature* sono strutturati nella forma <*value, likelihood*> dove *likelihood* può essere interpretata come la distribuzione di probabilità per un valore della *feature* dello specifico caso in oggetto¹⁴. Compattando la rappresentazione, possiamo considerare quindi gli annunci composti da una serie di triple <*feature, value, likelihood*>, interpretate come "la *feature* assume il valore *value* con probabilità *likelihood*". Le *feature* dell'annuncio possono inoltre essere obbligatorie: in questo caso sono vincolanti per la **selezione** degli elementi presentati all'utente finale (vedi colonna *constraints* che può assumere valori booleani: 0 e 1). Dunque la rappresentazione finale di ogni singola feature degli annunci è la seguente:

<*feature, value, likelihood, constraints*>

Si noti che le *feature* **disabilità** e **categoria protetta**, quando presenti, vengono sempre trattate dal sistema come vincolanti, quindi **obbligatorie**. Le feature descrittive possono essere tutte obbligatorie, dunque vincolanti e selettive, a seconda del tipo di formula di CBR applicata. Tutte le *feature*, quando non sono obbligatorie, contribuiscono invece all'ordinamento (**ranking**) degli elementi.

La tabella in Figura 15 mostra un esempio di **annuncio**, con le relative quadruple.

¹⁴ Cfr. Ardissono L., Goy A., Tailoring the Interaction with Users in Web Stores. User Model, User-Adaptive Interaction 10(4), pp. 251-303, 2000.

	Value	Likelihood	Constraints
idAnnuncio	34	-	-
<i>Descriptive Feature</i>			
Profilo	Barista	1	1
Gruppo	Personale di sala/bar	1	0
Tag	barista	0.333	0
	Barman	0.333	0
	Esperienza	0.333	0
<i>Preferential feature</i>			
Range età	24-30	1	1
Disabile	No	1	0
Categoria protetta	No	1	0
Titolo di studio	Licenza Media	1	1

Figura 15: Rappresentazione strutturata dell'annuncio.

Come si può notare la distribuzione di probabilità è sempre 1 quando il *value* è singolo, mentre si distribuisce uniformemente quando i value sono multipli secondo la seguente formula:

$$\text{Distribuzione di Probabilità (Pr)} = \frac{1}{\text{numero totale di value specificati}}$$

in altre parole, si assume che i valori alternativi di una certa feature, specificati dall'utente, siano ugualmente probabili nella descrizione del caso.

I dati nel modello utente si suddividono in:

- età
- disponibilità a lavorare all'estero
- lingua
- genere
- appartenenza alle categorie protette
- titolo di studio
- Il numero e la tipologia di profili selezionati
- I gruppi inferiti in base ai profili
- I tag inseriti

Profili, gruppi e tag, che possono contenere valori multipli, sono rappresentati secondo il seguente approccio:

- per ogni profilo/gruppo cui l'utente è associato il sistema automaticamente assegna una distribuzione di probabilità;
- i valori in una distribuzione di probabilità devono sempre sommare a 1;



- per ogni elemento di profilo/gruppo/tag i valori sono inizializzati ad una distribuzione di probabilità uniforme che rappresenta la stessa probabilità per ognuno di questi elementi, similamente a quanto si fa per gli annunci.

Seguono alcuni esempi chiarificatori:

Profilo

Per ogni elemento del profilo

Distribuzione di Probabilità (Pr) = 1/numero totale di profili specificati

Esempio: profiliUtente(p1,p2,p3)

Distribuzione di probabilità = 1/3

Pr(p1 = 0,333)

Pr(p2 = 0,333)

Pr(p3 = 0,333)

Esempio: profiliUtente(p1,p2)

Distribuzione di probabilità = 1/2

Pr(p1 = 0,5)

Pr(p2 = 0,5)

Gruppo

Per ogni profilo specificato

1. Considera anche il suo gruppo di appartenenza e porta a 1 il valore di probabilità parziale del gruppo a cui appartiene
2. Se il gruppo è già stato specificato incrementa il suo valore di probabilità parziale di 1

Per ogni gruppo

1. dividi la somma dei valori di probabilità parziali ottenuti per il totale dei gruppi cui l'utente appartiene

Esempio: profiliUtente(p1, p2, p3)

p1 ∈ g1

p2 ∈ g2

p3 ∈ g2

Pr(g1 = 1/3) = 0.33333333

Pr(g2 = 2/3) = 0.66666667

Come nel caso degli annunci, le *feature* del modello utente sono strutturate nella forma $\langle \text{feature}, \text{value}, \text{likelihood} \rangle$, interpretate come “la *feature* assume il valore *value* con probabilità *likelihood*”. La tabella in Figura 16 mostra la descrizione del modello utente completo di un **candidato**, con le relative triple

Features	Value	Likelihood
idCandidato	121	0
<i>Descriptive Feature</i>		
Profilo	Barista	0,5
	Barman	0,5
Gruppo	Personale di sala/bar	1
Tag	giovane	0.25
	Bar	0.25
	Cocktails	0.25
	Inglese	0.25
<i>Preferential Feature</i>		
Range età	24-30	1
Disabile	No	1
Categoria protetta	No	1
Titolo di studio	Maturità	1

Figura 16: Rappresentazione strutturata del candidato nel modello utente.

I casi vengono recuperati dalle formule attraverso due passaggi:

- **selezione**, che avviene sulle feature obbligatorie dell’annuncio (sia descrittive che preferenziali) che sono vincolanti. Quando vengono suggeriti dei candidati in base agli annunci, se questi ultimi contengono feature obbligatorie verranno proposti solo i candidati che le possiedono. Viceversa, ai candidati verranno proposti solo annunci che eventualmente richiedono feature obbligatorie compatibili con quanto da loro esplicitamente inserito;
- **ranking**, che lavora sia sulle feature obbligatorie che su quelle non obbligatorie in base al tipo di formula di CBR applicata. Nelle formule che seguono verranno descritti le selezioni e le strategie di ranking adottate per ottimizzare il match tra annuncio e candidato nelle varie formule di matchmaking proposte.

Il match tra i tag è un meccanismo sofisticato, che occorre spiegare nel dettaglio. A seguito dall’analisi della folksonomia condotta sul tagset di IoLavoro di ottobre 2012, è emersa in maniera netta la tendenza degli utenti (e dei candidati in particolare) a utilizzare tag composti, spesso descrittivi delle professioni da loro svolte: “istruttore tennis”, “operatore call-center”, ecc. In molti casi, inoltre, vengono usati termini stranieri o neologismi, non presenti nei dizionari disponibili per l’analisi automatica dei tag.

Per questo motivo, il meccanismo di confronto (match) tra tag dell’annuncio e tag dei candidati, sviluppato per Lavoro Interactive 3.0, non si limita a un confronto letterale tra

i tag, ma si basa su un'analisi della struttura interna dei tag, volta a individuare le relazioni sintattiche e semantiche tra tag composti. Il modello di strutturazione dei tag utilizzato per confrontare tag semplici e composti tiene quindi conto di un insieme di fattori linguistici, che includono caratteristiche morfologiche (plurale, singolare, ecc.), semantiche (similarità di concetti espressi da lemmi diversi: "addetto" e "operatore") e sintattiche (similarità tra strutture diverse: "vendite addetto" e "addetto vendita").

Il modello presentato di seguito è stato sviluppato a partire dai dati: nonostante esso risenta quindi della composizione del set di dati utilizzato, esso è caratterizzato da un approccio che mira a distinguere la parte del modello dipendente dai dati (liste di lemmi semanticamente simili, lemmi irrilevanti per il match) dal modello generale di struttura dei tag, che può essere facilmente adattato a un dominio diverso, nell'ambito della ricerca di lavoro.

- **Tag set.** Un tag set è un insieme di tag $T = \{t_1, \dots, t_n\}$
- **Tag.** Ogni tag contiene più parole $t_i = \{p_1, \dots, p_n\}$
- **Tipi di tag.** I tag possono appartenere a due **tipi di tag**:
 - semplice = contiene una sola parola
 - multi-parola = contiene più parole
- **Tipi di match.** Si definiscono i seguenti tipi di match tra tag: lessicale e semantico
- **Operazioni.** Lemmatizzazione: forme flesse di sostantivi e aggettivi vengono trasformate nella forma non flessa.
- **Liste di parole.**
 - La blacklist è una lista domain-specific di parole che vengono ignorate nell'analisi (addetto, responsabile, ecc.)
 - Stop words. Parole che non vengono sottoposte a nessun tipo di analisi (preposizioni, particelle, ecc.)

t_x e t_y sono entrambi tag semplici, oppure t_x e t_y sono entrambi tag multi-parola,

1. t_x e t_y sono **uguali** = uguaglianza tra stringhe (si escludono i tag semplici che corrispondono a parole in blacklist)

"capo villaggio" è uguale a "capo villaggio"

"torneo" è uguale a "torneo"

"addetto" e "addetto" vengono esclusi

t_x e t_y sono entrambi tag semplici lemmatizzati

2. t_x e t_y sono **uguali-L** = uguaglianza tra stringhe lemmatizzate (si escludono i tag che corrispondono a parole in blacklist)

"torneo" è uguale-L a "tornei"

t_x e t_y sono entrambi tag multi-parola (con tutte le parole lemmatizzate):

3. t_x e t_y sono **quasi uguali - AND** = esiste un ordine delle parole in t_x e t_y tale per

cui t_x è **uguali-l** a t_y

“direttore di tornei” è quasi uguale a “direttore tornei”, a “direttore torneo” e a “tornei direttore”

4. t_x e t_y sono **quasi uguali - OR** = una volta eliminate le *stop words* e le parole

in *blacklist*, vale la relazione di **uguale-l** tra almeno una parola di t_x e una parola di t_y

“direttore di tornei” è quasi uguale a “tornei organizzatore” e a “organizzatore torneo”

5. t_x e t_y sono **correlati-s** = entrambi i tag t_x e t_y sono composti dallo schema

$p_1 + p_2$

dove:

- p_1 è la prima parola di t_x e t_y

e p_1 appartiene alla *blacklist* oppure

appartiene alla lista dei nomi generici

AND

- p_2 è una parola p'_2 in t_x e p''_2

e p'_2 e p''_2 sono **semanticamente correlate**

(lista delle parole semanticamente correlate)

“istruttore sportivo” è correlato-s a “istruttore di tennis” e a “istruttore sport”

“addetto mensa” è correlato-s a “addetto ristorazione”

La lista dei nomi generici è un lista di sostantivi che acquistano un significato specifico solo se accompagnati dalla qualificazione data a un altro nome o aggettivo in un tag multi-parola (es. “istruttore”, “animatore”, “maestro”, ecc.).

La lista delle parole semanticamente correlate è una lista compilata sulla base dei tag multi-parola presenti nel corpus IoLavoro, costituita da insieme di lemmi (nomi e aggettivi) semanticamente correlati tra loro. Permette di attribuisce una relazione semantica generica (correlato-s) alle parole dello stesso insieme qualora appaiano in correlazione con un nome generico. Per es. {mensa, ristorante, ristorazione}, {sport, sportivo, tennis, nuoto, sci} nelle locuzioni “addetto mensa” e “addetto ristorante” o “maestro sport” e “maestro nuoto”.

t_x è un tag semplice (lemmatizzato) e t_y è un tag multi-parola (lemmatizzato), o viceversa

6. t_x e t_y sono in una relazione di tipo **contiene** = t_x è **uguale** a una parola contenuta in t_x oppure t_y è **uguale** a una parola contenuta in t_x

(escluse le parole in *blacklist*)

“direttore di tornei” è in una relazione di tipo contiene con “tornei”, “torneo” e “direttore”

t_x e t_w sono due tag semplici (lemmatizzati) e t_y è un tag multi-parola (lemmatizzati):

7. t_y è in una relazione di tipo **contiene-uguale** con t_x e t_w = la concatenazione di t_x e t_w (o di t_w e t_x)
è **uguale** a t_y
(includere parole in *blacklist*)

il tag complesso “addetto mensa” è in una relazione di tipo contiene-uguale con i due tag semplici “addetto” e “mensa”

Consideriamo due tagset: A e B

$A = \{t_{a_1}, \dots, t_{a_n}\}$

$B = \{t_{b_1}, \dots, t_{b_n}\}$

1. Per tutti i tag composti di A e B, si verifica se esiste una relazione di tipo 7 (contiene-uguale) tra il tag composto t_{a_i} , e una coppia di tag semplici (t_{a_j}, t_{a_k}) dell'altro insieme.
2. si formano tutte le coppie di tag dell'insieme $A \times B$ (prodotto cartesiano). Per ogni match di tipo 7 ottenuto al passo precedente, sono scartate dal prodotto cartesiano le coppie formate dal tag complesso e da ciascuno dei tag semplici (t_{a_i}, t_{a_j} e t_{a_i}, t_{a_k})
3. per ogni coppia,
 - a. Se i tag appartengono allo stesso tipo (entrambi semplici o entrambi multi-parola) si valuta se esiste un match di tipo 1.
 - b. Se non esiste un match di tipo 1 e i tag sono semplici, si valuta la presenza di un match di tipo 2
 - c. Se non esiste un match di tipo 1 e i tag sono multi parola: si valuta la presenza di un match di tipo 3
 - d. Se non esiste un match di tipo 1 e di tipo 3 e i tag sono multi parola: si valuta la presenza di un match di tipo 5
 - e. Se non esiste un match di tipo 1, di tipo 3 e di tipo 5 e i tag sono multi parola: si valuta la presenza di un match di tipo 4
 - f. Se i tag non appartengono allo stesso tipo, si valuta la presenza di un match di tipo 6.

Si considerano due tipi di relazioni semantiche tra i tag semplici:

- **Sinonimia:** t_1 e t_2 appartengono allo stesso synset

- **Iperonimia:** t_1 è iperonimo (termine più generale) di t_2 (t_1 appartiene a un synset *più generale* di uno dei synset a cui appartiene t_2)

Per ogni lista di tag:

1. Si suddividono i tag composti nelle loro componenti e li uniscono ai tag semplici, ottenendo una lista di tag semplici
2. Si lemmatizzano tutti i tag
3. Si escludono i tag che sono in *blacklist*
4. Si formano tutte le coppie di tag dell'insieme $A \times B$ (prodotto cartesiano)
5. Per ogni coppia di tag semplici lemmatizzati:

se sono uguali non c'è alcun match

altrimenti, verifica se t_1 e t_2 appartengono allo stesso synset (sinonimia)

altrimenti, verifica se t_1 è iperonimo di t_2 oppure t_2 è iperonimo di t_1

Ad ogni tipo di match è associato un valore, secondo i criteri seguenti. L'assegnazione dei valori alle relazioni tra tag si basa su un criterio di similarità semantica tra i tag, descritta attraverso un insieme di features (uguaglianza *ortografica*, *lessicale*, *morfologica*, *sintattica* e *semantica*). L'uguaglianza su tutte le features corrisponde all'identità del lemma, le altre relazioni sono ottenute dall'uguaglianza rispetto a un sottoinsieme di features.

I valori sono stati quindi distribuiti su una scala da 1 a 0 secondo i seguenti criteri di precedenza tra le relazioni identificate:

Match lessicale:

- uguali (w_1): relazione di uguaglianza *ortografica*, *lessicale*, *sintattica*, *morfologica*. Esempio: "capo villaggio" è uguale a "capo villaggio", "torneo" è uguale a "torneo"
- uguali-l (w_2): relazione di uguaglianza *ortografica*, *lessicale*, *sintattica* (non morfologica) Esempio: "torneo" è uguale-L a "tornei"
- contiene uguale (w_3): relazione di uguaglianza *lessicale*, *sintattica* (non *ortografica* e *morfologica*) Esempio: "addetto mensa" è in una relazione di tipo contiene-uguale "addetto" e "mensa"
- quasi uguali – AND (w_4): relazione di uguaglianza *lessicale* (non *morfologica* e *ortografica*) tra *tutti* i costituenti di due sintagmi (non uguaglianza *sintattica* tra i sintagmi) Esempio: "direttore di tornei" è quasi uguale a "direttore tornei", a "direttore torneo" e a "tornei direttore"

- quasi uguali – OR (w_5): relazione di uguaglianza lessicale (non morfologica e ortografica) tra *un sottoinsieme* dei costituenti di due sintagmi (non uguaglianza sintattica tra i sintagmi)
Esempio: “direttore di tornei” è quasi uguale a “tornei organizzatore” e a “organizzatore torneo”
- correlati-s (w_6): relazione di uguaglianza *lessicale* (non morfologica e ortografica) tra *tutti* i costituenti di due sintagmi (non uguaglianza *sintattica* tra i sintagmi), dove l’uguaglianza lessicale per almeno un termine è sostituita da una relazione di *similarità semantica*
Esempio: “istruttore sportivo” è correlato-s a “istruttore di tennis” e a “istruttore sport”.
- contiene (w_7): relazione di uguaglianza *lessicale* (non morfologica e ortografica) tra *un singolo* costituente di due sintagmi o di un sintagma e un lemma (quindi non *sintattica*)
Esempio: “direttore di tornei” è in una relazione di tipo contiene con “tornei”, “torneo” e “direttore”

A scopo sperimentale, utilizziamo i seguenti pesi. I valori sono assegnati su una scala da 0 a 1, sulla quale il 1 corrisponde all’uguaglianza (“addetto sala” è uguale a “addetto sala”), 0 alla mancanza di relazione (“torneista” non ha alcuna relazione con “programmatore web”).

Dopo aver raggruppato le relazioni quasi uguali – AND, quasi uguali – OR e correlati-s (in quanto definite da differenze quantitative rispetto alla stessa composizione di features) le relazioni sono state distribuite sulla scala decimale da 0 a 1 distribuendo le relazioni uguali – AND, quasi uguali – OR e correlati-s tra i valori intorno a 0,6.

Match lessicale: $w_1 = 1$; $w_2 = 0,8$; $w_3 = 0,7$; $w_4 = 0,6$; $w_5 = 0,5$; $w_6 = 0,4$; $w_7 = 0,2$

Per il match semantico, i valori sono stati assegnati considerando la relazione di sinonimia equivalente alla relazione quasi uguali – AND e la relazione di iperonimia equivalente alla relazione contiene

- tra profilo dell’annuncio e tag utente: si trattano il nome del profilo come un tag composto.



- tra titolo dell'annuncio e tag utente: si tratta il titolo come un insieme di tag semplici.

A seguito della sperimentazione, il titolo dell'annuncio è stato rimosso dall'insieme dei tag usati per il confronto lato utente. La grande varietà di titoli (per numero di parole e tipologia) infatti rendeva poco affidabile il match.

L'algoritmo per il match lessicale descritto in precedenza si applica quindi sui due insiemi di tag seguenti (insiemi A e B che costituiscono l'input dell'algoritmo):

- lato utente: tag utente
- lato annuncio: tag annuncio, nome del profilo (non si eliminano le duplicazioni)

Prima della fase di match dei tag, il valore di similarità basata sui tag tra annuncio e candidato viene inizializzata a zero.

A ogni match trovato tra i due insiemi A e B (vedi algoritmo precedente), si incrementa tale valore del valore corrispondente al tipo di match trovato.

Il valore di similarità finale tra l'insieme dei tag dell'utente e l'insieme dei tag dell'annuncio è ottenuto confrontando il valore ottenuto con un insieme di soglie.

<u>Soglia</u>	<u>Valore finale</u>
>0	0.25
>1	0.5
>2	0.75
>3	1

Il valore che la similarità basata sui tag assumerà nell'ambito delle formule di matching dipende dal peso assegnato alla componente "tag" nel modello CBR.

Per esempio, se il valore di similarità ottenuto dopo il confronto con il sistema delle soglie è 0.5, il valore finale sarà $0.25 = 0.5 * 0.5$.



Esempio 1

Si considerino i due insiemi di tag seguenti (candidato 1231, annuncio 57)

A	B
ballerino	ANIMATORE
coreografo	SPORTIVO
scenografo	ISTRUTTORE TENNIS
capo animatore	
istruttore sportivo	
insegnante danza hip hop	
istruttore nuoto	
assistente bagnante	
baby e junior club	
intrattenitore	
presentatore.	

I match ottenuti sono i seguenti:

w7	w7	w6	w6
capo animatore	istruttore sportivo	istruttore tennis	istruttore tennis
ANIMATORE	sportivo	istruttore sportivo	istruttore nuoto
0.2	0.2	0.4	0.4

Il valore ottenuto sommando i valori associati ai singoli tipi di match è:

$$0.2 + 0.2 + 0.4 + 0.4 = 1.2$$

tale valore viene confrontato con il sistema di soglie. La soglia superata è 1, quindi il valore finale sarà 0.5

Esempio 2

Si considerino i due insiemi di tag seguenti (candidato 1231, annuncio 361)

A	B
ballerino	ANIMATORI TURISTICI
coreografo	ADDETTI MINICLUB
scenografo	TORNEISTI SPORTIVI
capo animatore	ISTRUTTORI FITNESS/BALLI
istruttore sportivo	ASSISTENTI BAGNANTI
insegnante danza hip hop	
istruttore nuoto	
assistente bagnante	
baby e junior club	
intrattenitore	
presentatore.	
istruttore sportivo	

I match ottenuti sono i seguenti:

w5	w6	w2	w6
ANIMATORI TURISTICI	ISTRUTTORI FITNESS/BALLI	ASSISTENTI BAGNANTI	ISTRUTTORI FITNESS/BALLI
capo animatore	istruttore sportivo	assistente bagnante	istruttore sportivo
0.5	0.4	0.8	0.4

Il valore ottenuto sommando i valori associati ai singoli tipi di match è:

$$0.5 + 0.4 + 0.8 + 0.4 = 2.1$$

tale valore viene confrontato con il sistema di soglie. La soglia superata è 2, quindi il valore finale sarà 0.75.

Esempio 3

Si considerino i due insiemi di tag seguenti (candidato 7408, annuncio 416)

A	B
programmatore software	PROGRAMMATORE JAVA
gestore database	PROGRAMMATORE DOT.NET
analista	WEB ORIENTED
	ANALISTA FUNZIONALE

w5	w5	w7
PROGRAMMATORE JAVA	PROGRAMMATORE DOT.NET	ANALISTA FUNZIONALE
programmatore software	programmatore software	analista
0.4	0.4	0.2

Il valore ottenuto sommando i valori associati ai singoli tipi di match è:

$$0.4 + 0.4 + 0.2 = 1$$

tale valore viene confrontato con il sistema di soglie. La soglia superata è 0.5, quindi il valore finale sarà 0.25

Dato un insieme di elementi *target* t_i (che possono essere la descrizione di un annuncio o di un candidato) e un insieme di *elementi* c_i (che possono essere viceversa la descrizione di un candidato o di un annuncio) che hanno superato la fase di selezione in cui si prendono in considerazione solo i vincoli obbligatori, si propone un ranking degli elementi sulla base della funzione *Similarity* (t, c), espressa con la seguente formula:

$$Similarity(t, c) = \frac{\sum_{i=1..n} w_i \times sim_i(t_i, c_i)}{\sum_{i=1..n} w_i}$$

dove t è il target case (ad esempio, un annuncio) e c sono i casi correnti (ad esempio, i candidati), di cui ne verranno restituiti solo i k che hanno maggiore similarità con il target case, ordinati in modo decrescente (dal più simile al meno simile). Sia t che c sono descritti attraverso una serie di triple, indicate genericamente come t_i e c_i , come visto sopra (le quadruple entrano in gioco solo a livello di selezione). Ad ogni calcolo di similarità tra le tuple dell'annuncio e quelle del candidato viene inoltre assegnato un peso, w , che può essere *encoded* oppure appreso durante il funzionamento del sistema. I

pesi normalmente riflettono l'importanza che la *feature* descritta nella tupla riveste nel calcolo della similarità e assumono un valore che varia da 0 a 1, dove 1 rappresenta il massimo dell'importanza. Nel nostro caso i pesi w_i assumono un valore diverso a seconda dell'importanza della *feature* nel processo di raccomandazione così come emerso nelle varie fasi di testing. Dunque nel nostro caso i pesi saranno *encoded* e non appresi in corso d'utilizzo.

I valori di similarità sono dunque mediati dai pesi introdotti. I valori dei pesi sono dettagliati nelle sottosezioni riguardanti le singole formule di match. La funzione $\text{sim}()$ utilizzata per il calcolo di similarità tra le tuple dell'annuncio e quelle del candidato è una funzione di similarità simmetrica usata per il calcolo della similarità tra profili, gruppi e tag di annunci/candidati¹⁵:

$$\text{sim}(t,c) = 1 - |l_t - l_c|$$

Dovendo calcolare la similarità tra una *feature* espressa in range 0-1 ($l_t=0.8$) del target case t e la stessa *feature* ($l_c=0.7$) del target case c, la funzione di similarità simmetrica agisce come segue

$$\text{sim}(t,c) = 1 - \text{ass}(0,8 - 0,7) = 1 - \text{ass}(0,1) = 0,9$$

ossia si calcola la differenza tra i valori delle due *feature* (0,8-0,7) e se ne prende il valore assoluto ($\text{ass}(0,1)$), il quale viene infine sottratto a 1 e il risultato è il valore della similarità (0,9) tra le *feature*, espresso in range 0-1.

Nel nostro caso la $\text{sim}()$ viene calcolata sulle distribuzioni di probabilità del target case (l_t) e del current case (l_c) relativamente alle *feature* di tutte le tuple che descrivono annuncio e candidato.

Nel caso specifico invece del calcolo della similarità tra le *preferential feature* dell'annuncio e le *feature* possedute dai candidati, la similarità viene calcolata come segue, per poter assegnare un peso maggiore ai candidati che possiedono tutte le *feature* preferenziali. Si noti che, come già detto, le *feature* categoria protetta e disabile quando presenti, sono sempre e solo obbligatorie. Dunque la formula di cui sotto si applica solo alle *feature* *range età* e *titolo di studio*, quando indicate come non obbligatorie.

F_{pat} = *feature preferenziali totali dell'annuncio*

$F_{pa[i]}$ = *singola feature preferenziale dell'annuncio*

$F_{pc[i]}$ = *singola feature del candidato*

F_{pct} = *feature preferenziali del candidato che matchano con l'annuncio*

$$\text{sim_feature_preferenziali}(F_{pat}, F_{pct}) = F_{pct} / F_{pat}$$

Per ogni *feature* preferenziale

1. se la *feature* preferenziale del candidato *matcha* con quella dell'annuncio

¹⁵ Smyth, 2007, Opera citata.



incrementa di 1 le feature preferenziali del candidato che matchano con l'annuncio (F_{pct})

2. la similarità tra le feature preferenziali di candidato e annuncio sarà data dai rapporti tra le feature preferenziali del candidato che matchano con l'annuncio (F_{pct}) e feature preferenziali totali dell'annuncio F_{pat}

Si vedano i seguenti esempi.

Se l'utente possiede delle feature preferenziali dell'annuncio il valore di similarità sarà uguale al rapporto tra il numero di feature matchanti tra annuncio e candidato e il numero delle feature preferenziali totali dell'annuncio. Dunque se tutte le feature matchano la similarità è uguale a 1.

$$F_{pat} = 3$$

$$F_{pct} = 3$$

$$sim_feature_preferenziali(F_{pat}, F_{pct}) = 3 / 3 = 1$$

Se invece l'utente, ad esempio, possiede 2 feature su 3, il valore di similarità tra feature preferenziali sarà uguale a $2/3$, quindi 0,666.

$$F_{pat} = 2$$

$$F_{pct} = 3$$

$$sim_feature_preferenziali(F_{pat}, F_{pct}) = 2 / 3 = 0,666$$

Poiché si hanno comportamenti diversi, si sono differenziate le strategie e di conseguenza la presentazione dei risultati per le aziende e i candidati, rispettivamente.

L'azienda non può cercare attivamente candidati per i suoi annunci. Tuttavia il sistema suggerisce proattivamente e interattivamente dei candidati all'azienda per ognuno dei suoi annunci. Infatti l'utente aziendale, quanto seleziona uno dei suoi annunci, vedrà un box di candidati suggeriti dal sistema. Tali candidati sono scelti tra coloro che non si sono ancora candidati di propria iniziativa.

Date le seguenti variabili:

p_t = likelihood value of the target profile (annuncio)
 p_c = likelihood value of the current profile (candidato)
 g_t = likelihood value of the target group (annuncio)
 g_c = likelihood value of the current group (candidato)
 FO_t = mandatory features of the ads
 FO_c = mandatory features owned by the user

e le seguenti formule:

$sim_profili(p_t, p_c) = (1 - |p_t, p_c|) * w$
 $sim_gruppi(g_t, g_c) = (1 - |g_t, g_c|) * w$
 $sim_tag(tag_t, tag_c) = match\ lessicale - semantico\ tra\ tag_t\ e\ tag_c * w$

Per ogni annuncio il sistema propone i k best candidati (k da definire sulla base dei test) per l'annuncio selezionati in base al profilo dell'annuncio e alle sue eventuali feature obbligatorie, e ordinati poi attraverso l'algoritmo di CBR proposto secondo la Formula A1, ossia quella ristretta, sia in selezione che in ranking, solo al profilo richiesto dall'annuncio e alle eventuali feature obbligatorie.

Una volta calcolati tutti i candidati che fanno match vengono proposti i primi k con valore di Similarity() totale più alto (i cosiddetti *top ranked*). L'utente azienda potrà eventualmente scorrere i successivi k best secondo un approccio simile alla paginazione.

Ecco come funziona la formula A1:

- **Selezione:** profilo dell'annuncio ed eventuali feature obbligatorie

p_t è uno dei p_c
Per ogni FO_t , $FO_t == FO_c$

- **Ranking:** similarità tra profili di annuncio e candidati

$sim_profili(p_t, p_c) * w$
if $sim_profili(p_t, p_c) > 0,5 \rightarrow w=1$
else $w=0,8$

N.B. Si noti che la formula promuove i candidati che hanno una focalizzazione forte sul profilo richiesto. E' stata inserita una soglia di similarità pari a 0,5. Tale valore è stato introdotto per premiare i candidati molto focalizzati, ipotesi che è stata poi confermata nella prima fase di testing.

I suggerimenti che scaturiscono dall'ampliamento della ricerca sono quelli che derivano dalla Formula A2, ossia quella che nella selezione è allargata a gruppi e tag (e ovviamente eventuali feature obbligatorie), mentre nel ranking tiene conto della similarità tra profili, gruppi e tag (entrambi esclusi da A1). Anche in questo caso vengono proposti i k suggerimenti *top ranked* e con possibilità di paginazione.

- **Selezione:** eventuali feature obbligatorie, gruppi, tag

Per ogni FO_t , $FO_t == FO_c$
 g_t è uno dei g_c
Intersezione tag_t e tag_c non vuota

- **Ranking** sulla similarità tra profili, gruppi, tag

$sim_profili(p_t, p_c) * w$
 $if\ sim_profili(p_t, p_c) > 0,5 \rightarrow w=1, else\ w=0,8$
N.B. peso maggiore a candidati maggiormente focalizzati
 $sim_gruppi(g_t, g_c) * w$ ($w=1$)
 $sim_tag(tag_t, tag_c) * w$ (w dipendente dal tipo di match, max 0,5)

Questi suggerimenti sono stati ideati per proporre dei risultati non perfettamente matchanti, dunque non nei *top ranked*, ma in qualche modo correlati alla ricerca, con l'obiettivo di allargare l'orizzonte delle possibilità offerte. In questa funzionalità intelligente (cosiddetta STUPISCIMI!) tra la rosa dei candidati che matchano con l'annuncio (computati secondo il suggerimento A2) vengono selezionati i k candidati ($k=10$) che hanno un'alta similarità alta con l'annuncio, ma sono dissimili tra di loro. Per il calcolo della dissimilarità tra gli utenti si utilizza l'indice di Tanimoto.

Ricordiamo che la **distanza di Tanimoto** è una variante dell'indice di Jaccard (vedi prima) per la similarità tra vettori. Nel nostro caso, ogni utente è visto come descritto da un vettore di profili, $u = [p_1, \dots, p_n]$ e dunque viene calcolata la similarità con tutti gli altri vettori di profili che descrivono gli utenti utilizzando il rapporto tra l'intersezione dei vettori diviso la loro unione.

Il primo utente proposto è sempre quello che era al primo posto nel suggerimento A2, mentre il secondo sarà il primo maggiormente dissimile dall'elemento in posizione 1 (con indice di Tanimoto < 0.5 calcolato sui profili) all'interno della lista ordinata costruita secondo il suggerimento A2. Il terzo sarà il primo maggiormente dissimile dall'elemento in posizione 2 tra quelli successivi nei *top ranked* e così via.

Tale suggerimento è ispirato alla *bounded greedy selection strategy*¹⁶. L'algoritmo

¹⁶ Cfr. "Then the bounded greedy selection strategy proposed by [...] offers a way to improve diversity, while at the same time maintaining target query similarity; see also [...]. This strategy incrementally builds

funziona in questo modo

1. Nella rosa di candidati A2 si selezionano i migliori - quindi quelli con un alto valore di `similarity()`- k ($k=10$) ma dissimili tra di loro (con indice di Tanimoto $< 0,5$ in range 0-1 tra i candidati che hanno almeno un profilo in comune).
2. Per il calcolo della dissimilarità tra gli utenti si utilizza la distanza di Tanimoto.
3. La distanza di Tanimoto:
 - a. utente come vettore di profili, $u = [p_1 \dots p_n]$
 - b. similarità con tutti gli altri vettori di profili = rapporto tra l'intersezione dei vettori diviso la loro unione.

L'utente trova i suggerimenti sia accede alla sua myPage sia quando fa delle ricerche (i risultati sono sempre ordinati secondo un ranking).

Il sistema lavora in maniera allargata considerando l'algoritmo di similarità specificato nella Formula C1, che è una variante della formula A2, ossia quella basata sull'uso ampliato della similarità, che considera gruppi, tag e feature obbligatorie nella selezione e profili, gruppi e tag nel ranking. All'utente vengono proposti nel box i k annunci con similarità più alta rispetto al suo profilo. L'utente può navigare nei successivi *top ranked* usando la paginazione.

- **Selezione:** eventuali feature obbligatorie, gruppi, tag

*Per ogni FO_t, FO_c $FO_t == FO_c$
 g_t è uno dei g_c
Intersezione tag_t e tag_c non vuota*

- **Ranking** sulla similarità tra profili, gruppi, tag

*$sim_profili(p_t, p_c) * w$
 $if\ sim_profili > 0,5 \rightarrow w=1, else\ w=0,8$
N.B. peso maggiore a candidati maggiormente focalizzati
 $sim_gruppi(g_t, g_c) * w$ ($w=1$)
 $sim_tag(tag_t, tag_c) * w$ (w dipendente dal tipo di match, max 0,5)*

L'utente può filtrare i suo *top ranked* (calcolati secondo C1) specificando uno dei suoi profili.

a diverse retrieval set of k cases, R , by starting from a set of the b_k most similar cases to the target query. During each step, the remaining cases are ordered according to their quality with the highest quality case added to R . (...) The quality of a case c is proportional to the similarity between c and the current target t , and to the diversity of c relative to those cases so far selected, $R = \{r_1, \dots, r_m\}$," Smyth, 2007, opera citata.

Quindi se l'utente ha specificato i profili p_1 e p_2 , può chiedere di vedere solo gli annunci relativi a p_1 o a p_2 , rispettivamente, che verranno presentati ordinati secondo i punteggi calcolati precedentemente in C1.

Come nel caso di A3, questi suggerimenti sono stati ideati per proporre dei risultati non perfettamente matchanti, dunque non nei *top ranked*, ma in qualche modo correlati alla ricerca, con l'obiettivo di allargare l'orizzonte delle possibilità offerte. Anche questa modalità STUPISCIMI è mirata a favorire la dissimilarità tra i risultati. Tuttavia il calcolo della distanza di Tanimoto tra gli annunci potrebbe dare problemi sul numero di annunci veramente dissimili (ad es. vi sarebbe troppa poca differenza tra gli annunci se l'utente avesse pochi profili). Quindi si è pensato di usare una bounded random selection, strategia che seleziona k casi scelti casualmente nella lista dei *top ranked* raggruppati secondo delle fasce (es. seleziona casualmente un candidato con similarità tra 1 e 0.9, e poi casualmente uno con similarità tra 0.9 e 0.8, e così via).

L'algoritmo funziona in questo modo:

1. Si selezionano k casi ($k=10$) scelti casualmente nella lista dei *top ranked* in C1 raggruppati secondo delle fasce
 - i. Selezione casuale di un annuncio con similarità in $\langle 0.9, 1 \rangle$
 - ii. Selezione casuale di un annuncio con similarità in $\langle 0.8, 0.9 \rangle$
2. ... fino a trovare 10 annunci diversi

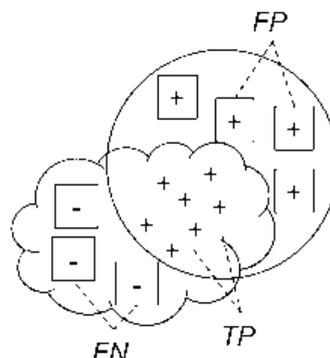
Dal momento che le agenzie regionali del lavoro (APL, ALL e Regione VdA), committenti del progetto Lavoro Interactive, hanno una grande esperienza nella gestione dell'incontro tra domanda e offerta di lavoro stagionale è stato loro richiesto di partecipare alla valutazione come esperti di dominio.

Il loro compito nella valutazione è stato quello di simulare le azioni degli utenti reali e fornire un giudizio binario sulla bontà delle raccomandazioni proposte, formula per formula. Le metriche principali per valutare la bontà o meno dei suggerimenti proposti da un sistema, e giudicati in maniera binaria (es. mi piace, non mi piace), sono la *precision* e la *recall*. Queste metriche derivano dalle valutazioni fatte nei sistemi di *information retrieval* in quanto il problema è molto simile al nostro: da un insieme molto grande di contenuti un sotto-insieme di contenuti rilevanti per l'utente deve essere estratto.

La **precision** si riferisce al grado di accuratezza del processo di selezione di un recommender system¹⁷. La precision è il rapporto tra i contenuti rilevanti per l'utente e i contenuti presentati all'utente stesso. La precision è data dalla formula

$$\frac{TP}{TP + FP}$$

dove TP sono i **veri positivi**, ossia i contenuti rilevanti che sono anche presentanti all'utente, FP sono i **falsi positivi**, ossia i contenuti che sono suggeriti all'utente ma che lui non considera rilevanti. La precision viene giudicata buona quando è superiore al 70%.



La **recall** è il rapporto tra i contenuti rilevanti per l'utente e tutti i contenuti della collezione, compresi quelli che non sono stati suggeriti all'utente. La recall è data dalla formula

$$\frac{TP}{TP + FN}$$

dove TP sono i **veri positivi**, ossia i contenuti rilevanti che sono anche presentanti all'utente e FN sono i **falsi negativi**, ossia i contenuti che non sono stati suggeriti all'utente ma che lui considera rilevanti. Anche la recall viene giudicata buona quando è superiore al 70%.

Quindi la precision indica quanto un sistema che privilegia il ranking è accurato, mentre la recall indica quando un sistema è accurato nel trovare l'informazione che interessa l'utente. Nel caso del nostro sistema di machmaking, il dato per noi più importante è quello riguardante la precision. Tuttavia se calcoliamo anche la recall abbiamo una stima di quanti contenuti rilevanti per l'utente stiamo escludendo. Si noti che, nella tabella in Figura 17, la **recall** è calcolata come il rapporto tra:

$$\frac{\text{contenuti rilevanti per l'utente nei } k \text{ best}}{\text{tutti i contenuti votati come rilevanti}}$$

¹⁷ Salton G., McGill M., Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1984.



formula	totali su POSIZIONE 10		formula	positivi su 10 (precision)		Recall	negativi su 10	
A1	221	22,48%	A1	174	78,73%	15,70%	47	21,27%
A2	108	10,99%	A2	47	43,52%	27,65%	61	56,48%
A3	17	1,73%	A3	4	23,53%	100,00%	13	76,47%
C1	226	22,99%	C1	165	73,01%	67,07%	61	26,99%
C2	131	13,33%	C2	100	76,34%	100,00%	31	23,66%
C3	280	28,48%	C3	81	28,93%	100,00%	199	71,07%
% SU 10	983	100,00%		571	58,09%	33,41%	412	41,91%

formula	totali su POSIZIONE 20		formula	positivi su 20 (precision)		Recall	negativi su 20	
A1	365	27,80%	A1	298	81,64%	26,90%	67	18,36%
A2	213	16,22%	A2	80	37,56%	47,06%	133	62,44%
A3	17	1,29%	A3	4	23,53%	100,00%	13	76,47%
C1	299	22,77%	C1	219	73,24%	89,02%	80	26,76%
C2	132	10,05%	C2	100	75,76%	100,00%	32	24,24%
C3	287	21,86%	C3	81	28,22%	100,00%	206	71,78%
% SU 20	1313	100,00%		782	59,56%	45,76%	531	40,44%

Formula	totali su POSIZIONE 30		formula	positivi su 30 (precision)		Recall	negativi su 30	
A1	489	31,59%	A1	393	80,37%	35,47%	96	19,63%
A2	311	20,09%	A2	113	36,33%	66,47%	198	63,67%
A3	17	1,10%	A3	4	23,53%	100,00%	13	76,47%
C1	312	20,16%	C1	231	74,04%	93,90%	81	25,96%
C2	132	8,53%	C2	100	75,76%	100,00%	32	24,24%
C3	287	18,54%	C3	81	28,22%	100,00%	206	71,78%
% SU 30	1548	100,00%		922	59,56%	53,95%	626	40,44%



Formula	totali		totale positivi (precision)		totale negativi	
A1	2036	63,31%	1108	54,42%	928	45,58%
A2	432	13,43%	170	39,35%	262	60,65%
A3	17	0,53%	4	23,53%	13	76,47%
C1	326	10,14%	246	75,46%	80	24,54%
C2	133	4,14%	100	75,19%	33	24,81%
C3	272	8,46%	81	29,78%	191	70,22%
	3216	100,00%	1709	53,14%	1507	46,86%

Figura 17: Dati di precision e recall della prima fase di valutazione

I dati sopra riportati sono stati analizzati nel dettaglio e hanno contribuito alle modifiche delle formule che verranno dettagliate nelle pagine successive (vedi formule seconda versione). Riassumiamo qui di seguito le principali osservazioni emerse dopo la prima fase di valutazione:

- A1: la formula è ben performante e addirittura migliora al crescere dei risultati, ma è deludente per quanto riguarda la recall, ossia esclude ancora molti risultati positivi dai *top ranked*;
- A2: i risultati generati dalla formula sono stati valutati in maniera molto deludente;
- A3: la formula ha prodotto risultati estraneamente deludenti;
- C1-C2: entrambe le formule hanno ottenuto buoni risultati, in termini sia di precision sia di recall;
- C3: la formula è stata valutata in maniera estremamente deludente.

Dati I risultati ottenuti, abbiamo modificato le formule. Nel seguito, si assumeranno sempre le seguenti variabili

p_t = likelihood value of the target profile (annuncio)
 p_c = likelihood value of the current profile (candidato)
 g_t = likelihood value of the target group (annuncio)
 g_c = likelihood value of the current group (candidato)
 FO_t = mandatory features of the ads
 FO_c = mandatory features owned by the user
 F_{pat} = total preferential features of the job advertisement
 F_{pct} = total matching features of the job candidate

e le seguenti formule di base

$sim_profili(p_t, p_c) = (1 - |p_t, p_c|) * w$
 $sim_gruppi(g_t, g_c) = (1 - |g_t, g_c|) * w$
 $sim_feature_preferenziali(F_{pat}, F_{pct}) = F_{pct} / F_{pat}$
 $sim_tag(tag_t, tag_c) = match\ lessicale\ o\ semantico\ tra\ tag_t\ e\ tag_c * w$

riproposte qui, per facilitare il lavoro dei programmatori, che usano questo documento come riferimento.

Per ogni annuncio il sistema propone i *k best* candidati (k da definire sulla base dei test) per l'annuncio selezionati in base al profilo dell'annuncio e alle sue eventuali feature obbligatorie, e ordinati poi attraverso l'algoritmo di CBR proposto secondo la nuova Formula A1, che ora considera sia il profilo che il gruppo di candidati e annuncio, e le eventuali feature preferenziali.



Una volta calcolati tutti i candidati che fanno match vengono proposti i primi k con valore di `Similarity()` totale più alto (i cosiddetti *top ranked*). L'utente azienda può eventualmente scorrere i successivi k best secondo un approccio simile alla paginazione.

- **Selezione:** profilo e eventuali FO (uguale alla prima versione della formula)

p_t è uno dei p_c
Per ogni FO_t , $FO_t == FO_c$

- **Ranking:** similarità tra profili di annuncio e candidati (uguale alla prima versione della formula)

$sim_profili(p_t, p_c) * w$
if $sim_profili(p_t, p_c) > 0,5 \rightarrow w=1$, else $w=0,8$

Si continua ad assegnare un peso maggiore a candidati maggiormente focalizzati.

Le novità principali rispetto alla prima versione della formula sono nelle strategie di ranking, dove vengono tenuti in conto sia il gruppo dell'annuncio che le eventuali feature preferenziali.

- **Ranking:** similarità tra gruppo dell'annuncio e gruppi del candidato

$sim_gruppi(g_t, g_c) * w$ ($w=1$)

Dai risultati della valutazione si è visto, infatti, che i candidati con maggiore focalizzazione sul gruppo hanno ottenuto dei feedback migliori

- **Ranking:** candidato con feature preferenziali dell'annuncio

$sim_feature_preferenziali(F_pat, F_pct) * w$ ($w=1$)

Dai risultati della valutazione è emerso quanto sia importante che i candidati possiedano il maggior numero di feature non obbligatorie (preferenziali) richieste dall'annuncio

Esempio di applicazione di A1 (si noti la non partecipazione dei tag al calcolo)

(annuncio)
profilo[barista: 1]
gruppo[Personale di sala/bar: 1]
tag[bar: 0,5, barista: 0,5]
features[eta: 18-24, titolo_studio: maturità]

(utente)
profilo[barista: 0,5, magazziniere 0,5]
gruppo[Personale di sala/bar: 0,5, logistica: 0,5]



tag[bar: 0,333, magazzino: 0,333, cameriere: 0,333]
features[eta: 25-34, titolo_studio: maturità]

A1

$$\text{sim_profili} = (1 - \text{ass}(1-0,5)) * 0,8 = 0,5 * 0,8 = 0,4$$

$$\text{sim_gruppi} = (1 - \text{ass}(1-0,5)) * 1 = 0,5 * 1 = 0,5$$

$$\text{sim_feature_preferenziali} = (1/2) * 1 = 0,5 * 1 = 0,5$$

$$\text{sim_TOTALE} = 0,4 + 0,5 + 0,5 / 0,8 + 1 + 1 = 1,4 / 2,8 = 0,5$$

I suggerimenti che scaturiscono dall'ampliamento della ricerca sono quelli che derivano dalla nuova Formula A2 che nella selezione tiene ora conto solo del gruppo dell'annuncio e delle eventuali feature obbligatorie, mentre nel ranking considera invece la similarità tra profili, gruppi, tag e le eventuali feature preferenziali. Il ranking combina questi elementi con pesi diversi, il cui valore è stato induttivamente influenzato dai risultati della valutazione, nella formula del CBR, come sotto dettagliato.

Anche in questo caso vengono proposti i k suggerimenti *top ranked* e con possibilità di paginazione.

- **Selezione:** eventuali feature obbligatorie, gruppo

*Per ogni FO_t, FO_t == FO_c
g_t è uno dei g_c*

È stata **eliminata la selezione sui tag** comuni tra annuncio e candidati (l'allargamento ai tag includeva troppi candidati giudicati irrilevanti in fase di valutazione generando inoltre un numero eccessivo di risultati su cui calcolare il ranking).

- **Ranking** sulla similarità tra tag (uguale alla prima versione della formula)

$$\text{sim_tag}(\text{tag}_t, \text{tag}_c) * w \quad (w \text{ dipendente dal tipo di match, max } 0,5)$$

Le novità sono:

- **Ranking** sulla similarità tra profili

*If p_t è uno dei p_c allora sim_profili(p_t, p_c) = 1 (w=1)
else sim_profili(p_t, p_c) = 0*

N.B. dopo la prima valutazione, è stata decisa una polarizzazione sui candidati che possiedono almeno il profilo dell'annuncio, in quanto veniva ritenuta dai valutatori una condizione necessaria, anche se non sufficiente, per essere valutati al meglio)

- **Ranking** sulla similarità tra gruppi

$$\text{sim_gruppi}(g_t, g_c) * w \quad (w=0,25)$$



Dopo la prima valutazione, è stata diminuita la focalizzazione sul gruppo (diminuendone il peso) in quanto ritenuto meno rilevante delle altre componenti del ranking (profilo, tag, feature preferenziali). Il gruppo in A2 è fondamentale nella selezione, meno nel ranking

- **Ranking:** candidato con feature preferenziali dell'annuncio

$$\text{sim_feature_preferenziali}(F_pat, F_pct) * w \quad (w=0,75)$$

Dai risultati della valutazione è emerso quanto sia importante che i candidati possiedano il maggior numero di feature non obbligatorie (preferenziali) richieste dall'annuncio

Esempio di applicazione di A2

(annuncio)

profilo[barista: 1]
gruppo[Personale di sala/bar: 1]
tag[bar: 0,5, barista: 0,5]
features[eta: 18-24, titolo_studio: maturità]

(utente)

profilo[barista: 0,5, magazziniere 0,5]
gruppo[Personale di sala/bar: 0,5, logistica: 0,5]
tag[bar: 0,333, magazzino: 0,333, cameriere: 0,33]
features[eta: 25-34, titolo_studio: maturità]

A2

$$\begin{aligned} \text{sim_profili} &= 1 * 1 = 1 \\ \text{sim_gruppi} &= (1 - \text{ass}(1-0,5)) * 0,25 = 0,5 * 0,25 = 0,125 \\ \text{sim_feature_preferenziali} &= (1/2) * 0,75 = 0,5 * 0,75 = 0,375 \\ \text{sim_tag} &= 0,25 * 0,5 = 0,125 \\ \text{sim_TOTALE} &= 1 + 0,125 + 0,375 + 0,125 / 1 + 0,25 + 0,75 + 0,5 = 1.625 / 2,5 = 0,65 \end{aligned}$$

In questa funzionalità (cosiddetta STUPISCIMI!), tra la rosa dei candidati che matchano con l'annuncio (computati secondo il suggerimento A2) vengono selezionati i k candidati (k=10) che hanno un'alta similarità alta con l'annuncio, ma sono dissimili tra di loro. Per il calcolo della dissimilarità tra gli utenti si utilizza l'indice di Tanimoto.

Visti i risultati deludenti ottenuti in fase di prima valutazione l'algoritmo è stato modificato in maniera tale da includere una rosa più ristretta di candidati (non oltre i primi 100 classificati). Dunque si è pensato di fare in questo modo: il primo utente proposto è sempre quello che era al primo posto nel suggerimento A2, mentre il secondo sarà il primo maggiormente dissimile dall'elemento in posizione 1 (con indice di Tanimoto < 0.5 calcolato sui profili) nei successivi 10 all'interno della lista ordinata costruita secondo il suggerimento A2. Il terzo sarà il primo

maggiormente dissimile dall'elemento in posizione 2 tra i 10 successivi e così via. Tale suggerimento è ispirato alla *bounded greedy selection strategy*¹⁸.

L'algoritmo che riassume le considerazioni precedenti è il seguente:

1. primo candidato A2 = primo classificato A3 (Stupiscimi_corrente)
2. Considera i 10 elementi successivi alla posizione Stupiscimi_corrente che abbiano almeno un profilo in comune con Stupiscimi_corrente
3. Ordinare i 10 elementi dal meno simile al più simile a Stupiscimi_corrente secondo l'indice di similarità di Tanimoto sui profili
4. Il primo candidato (il meno simile) dei 10 diventa il nuovo Stupiscimi_corrente
5. Si riparte dal punto 2 con la posizione successiva a Stupiscimi_corrente fino a trovarne 10

L'utente trova i suggerimenti sia accede alla sua myPage sia quando fa delle ricerche (i risultati sono sempre ordinati secondo un ranking).

Il sistema lavora in maniera allargata considerando l'algoritmo di similarità specificato nella formula C1, che è una variante della formula A2, ossia quella basata sull'uso ampliato della similarità, e che seleziona in base a gruppi e feature obbligatorie, e ordina in base a profili, gruppi, tag e feature preferenziali. All'utente vengono proposti nel box i k annunci con similarità più alta rispetto al suo profilo. L'utente può navigare nei successivi *top ranked* usando la paginazione.

- **Selezione:** eventuali feature obbligatorie e gruppi specificati dal candidato

*Per ogni FO_t, FO_t == FO_c
g_t è uno dei g_c*

È stata **eliminata la selezione sui tag** comuni tra annuncio e candidati (l'allargamento ai tag includeva troppi candidati giudicati irrilevanti in fase di valutazione generando inoltre un numero eccessivo di risultati su cui calcolare il ranking)

- **Ranking** sulla similarità tra tag

*sim_tag(tag_t, tag_c) * w (w dipendente dal tipo di match, max 0,5)*

- **Ranking** sulla similarità tra profili

*sim_profili(p_t, p_c) * w
if sim_profili > 0,5 then w=1, else w=0,8*

¹⁸ Smyth, 2007, opera citata.



Le novità sul ranking sono:

- **Ranking** sulla similarità tra gruppi

$$\text{sim_gruppi}(g_t, g_c) * w \quad (w=0,25)$$

Dopo la prima valutazione, è stata diminuita la focalizzazione sul gruppo in fase di ranking in quanto giudicato dai tester meno rilevante

- **Ranking**: gli annunci che richiedono feature che matchano con quelle del candidato

$$\text{sim_feature_preferenziali}(F_pat, F_pct) * w \quad (w=0,75)$$

Dai risultati della valutazione è emerso quanto sia importante che i candidati possiedano il maggior numero di feature non obbligatorie (preferenziali) matchanti con gli annunci proposti.

Esempio di applicazione di C1

(annuncio)

profilo[barista: 1]
gruppo[Personale di sala/bar: 1]
tag[bar: 0,5, barista: 05]
features[eta: 18-24, titolo_studio: maturità]

(utente)

profilo[barista: 0,5, magazziniere 0.5]
gruppo[Personale di sala/bar: 0.5, logistica: 0.5]
tag[bar: 0,333, magazzino: 0, 333, cameriere: 0,3333]
features[eta: 25-34, titolo_studio: maturità]

C1

$$\begin{aligned} \text{sim_profili} &= (1 - \text{ass}(1-0,5)) * 1 = 0,5 * 1 = 0,5 \\ \text{sim_gruppi} &= (1 - \text{ass}(1-0,5)) * 0,25 = 0,5 * 0,25 = 0,125 \\ \text{sim_feature_preferenziali} &= (1/2) * 0,75 = 0,5 * 0,75 = 0,375 \\ \text{sim_TAG} &= 0,25 * 0,5 = 0,125 \\ \text{sim_TOTALE} &= 0,5 + 0,125 + 0,375 + 0,125 / 1 + 0,25 + 0,75 + 0,5 = 1,125 / 2,5 = 0,45 \end{aligned}$$

L'utente può filtrare i suo *top ranked* (calcolati secondo C1) specificando uno dei suoi profili. Quindi se l'utente ha specificato i profili p1 e p2, può chiedere di vedere solo gli annunci relativi a p1 o a p2, rispettivamente, che verranno presentanti ordinati secondo i punteggi calcolati precedentemente in C1.



Visti i deludenti risultati della prima fase di valutazione, si è deciso di abbandonare l'approccio casuale implementato, e realizzare uno stupiscimi simile ad A3.

Ecco il calcolo:

1. primo candidato classificato in C1 = primo classificato in C3 (Stupiscimi_corrente)
2. Considerare i 10 elementi successivi alla posizione Stupiscimi_corrente
3. Ordinare i 10 elementi dal meno simile al più simile a Stupiscimi_corrente secondo l'indice di similarità di Tanimoto sui profili
4. il primo candidato (il meno simile) dei 10 diventa il nuovo Stupiscimi_corrente
5. si riparte dal punto 2 con la posizione successiva a Stupiscimi_corrente fino a trovarne 10

L'approccio di valutazione usato in seconda fase è stato il medesimo della prima, anche per poter confrontare i risultati. Dunque APL, ALL e Regione VdA hanno simulato le azioni degli utenti reali e hanno fornito un giudizio binario sulla bontà delle raccomandazioni proposte, formula per formula.

Sui risultati così ottenuti sono stati calcolati *precision* e *recall* (Figura 18) e abbiamo calcolato la differenza delle performance con i risultati ottenuti con la prima versione (Figura 19).

formula	totali su POSIZIONE 10		formula	positivi su 10 (precision)		recall	negativi su 10	
A1	285	24,34%	A1	228	80,00%	26,89%	57	20,00%
A2	214	18,27%	A2	172	80,37%	11,85%	42	19,63%
A3	120	10,25%	A3	69	57,50%	100,00%	51	42,50%
C1	244	20,84%	C1	216	88,52%	13,57%	28	11,48%
C2	116	9,91%	C2	113	97,41%	100,00%	3	2,59%
C3	192	16,40%	C3	133	69,27%	100,00%	59	30,73%
% SU 10	1171	100,00%		931	79,50%		240	20,50%

formula	totali su POSIZIONE 20		formula	positivi su 20 (precision)		recall	negativi su 20	
A1	400	22,26%	A1	324	81,00%	38,21%	76	19,00%
A2	431	23,98%	A2	304	70,53%	20,95%	127	29,47%
A3	120	6,68%	A3	69	57,50%	100,00%	51	42,50%
C1	491	27,32%	C1	401	81,67%	25,19%	90	18,33%
C2	116	6,46%	C2	113	97,41%	100,00%	3	2,59%
C3	239	13,30%	C3	133	55,65%	100,00%	106	44,35%
% SU 20	1797	100,00%		1344	74,79%		453	25,21%

formula	totali su POSIZIONE 30		formula	positivi su 30 (precision)		recall	negativi su 30	
A1	470	20,26%	A1	385	81,91%	45,40%	85	18,09%
A2	647	27,89%	A2	430	66,46%	29,63%	217	33,54%
A3	120	5,17%	A3	69	57,50%	100,00%	51	42,50%
C1	728	31,38%	C1	542	74,45%	34,05%	186	25,55%
C2	116	5,00%	C2	113	97,41%	100,00%	3	2,59%
C3	239	10,30%	C3	133	55,65%	100,00%	106	44,35%
% SU 30	2320	100,00%		1672	72,07%		648	27,93%

formula	totali		totale positivi (precision)		totale negativi	
A1	938	8,53%	848	90,41%	90	9,59%
A2	3860	35,11%	1451	37,59%	2409	62,41%
A3	120	1,09%	69	57,50%	51	42,50%
C1	5721	52,04%	1592	27,83%	4129	72,17%
C2	116	1,06%	113	97,41%	3	2,59%
C3	239	2,17%	133	55,65%	106	44,35%
	10994	100,00%	4206	38,26%	6788	61,74%

Figura 18: Dati di precision e recall della seconda fase di valutazione

formula	INCREMENTI/DECREMENTI	
	precision	Recall
A1	1,27%	11,18%
A2	36,86%	-15,79%
A3	33,97%	0,00%
C1	15,52%	-53,51%
C2	21,59%	-1,77%
C3	36,82%	-19,55%

formula	INCREMENTI/DECREMENTI	
	precision	Recall
A1	-0,64%	11,31%
A2	32,97%	-26,11%
A3	33,97%	0,00%
C1	8,43%	-63,84%
C2	21,66%	0,00%
C3	27,43%	0,00%

formula	INCREMENTI/DECREMENTI	
	precision	Recall
A1	1,55%	9,93%
A2	30,13%	-36,84%
A3	33,97%	0,00%
C1	0,41%	-59,86%
C2	21,66%	0,00%
C3	27,43%	0,00%

Figura 19: Confronti tra i risultati della prima e della seconda fase di valutazione.

Dalla Figura 19, che fa vedere i confronti tra i risultati, se ci si concentra sui dati della nuova valutazione, si può notare che la recall migliora al crescere delle posizioni di ranking considerate. Quindi l'aumento dei k best genera anche un aumento dei contenuti rilevanti per l'utente.

Già con k=20 le formule C2, A3 e C3 raggiungono il 100% di recall, e questo è dovuto al fatto che queste formule restituiscono sempre pochi risultati. La formula A1 è quella che viene valutata meglio in termini di precision e recall su k=30, seguita da C1 e infine da A2.

Guardando il confronto con i dati passati (Figura 19) si nota che A2 è molto migliorata in termini di precision a scapito della recall (fatto positivo), mentre C1 è rimasta stabile nella precision ma è peggiorata nella recall. Dunque nella valutazione precedente i falsi negativi esclusi erano di meno. Se si considera però che nella seconda fase di valutazione C1 ha

ottenuto il 50% dei voti totali (positivi e negativi) non stupisce che la recall sia diminuita. Tuttavia C1, ossia la formula che suggerisce annunci ai candidati, non esaurisce nei suoi primi 30 i suoi buoni suggerimenti. Dunque si potrebbe pensare di cambiare la formula C3 (stupiscimi) al fine riuscire ad includere un maggior numero di buoni risultati esclusi. Ad esempio C3 potrebbe iniziare i suoi calcoli dall'elemento $z+1$ con $z = 30$.

Stesso discorso vale per A2 (che suggerisce in senso allargato i candidati alle aziende) che secondo i dati di precision si dovrebbe fermare a $k = 30$. Tuttavia il dato basso di recall indica che vengono esclusi molti buoni candidati che potrebbero essere recuperati modificando la formula A3 facendola iniziare da un primo elemento in posizione $z+1$ con $z = 30$.

Questi cambiamenti nelle formule A3 e C3 sicuramente peggioreranno la precision a scapito della recall. Dunque otterranno meno consensi generali dagli utenti, ma saranno in grado di includere dei risultati esclusi dalle altre formule, migliorando quindi la recall.

Infine, considerando che una buona precision è quella che raggiunge una performance vicina al 70%, si consiglia di fermare i top ranked per A1, A2, C1 ai primi 30 risultati ottenenti.

Lo studio di interfaccia per la presentazione dei risultati è riportato in Appendice. L'interfaccia include una fiera virtuale 3D, che propone una selezione degli annunci in ricerca libera e una visita che parte dall'alto degli stand di una fiera reale simulata per proporre accessi agli stand diretti con volo d'uccello o percorsi guidati a terra sulla base della ricerca effettuata.

Nell'interfaccia testuale, i risultati sono ordinati secondo quanto ogni annuncio è compatibile (matcha) con il modello utente. Oltre alle proposte selezionate e ordinate dal sistema, il candidato può fare ricerche tra gli annunci filtrando per nome dei profili cercati, località e per tag. Queste chiavi di ricerca possono essere congiunte (in AND) oppure anche fatte singolarmente.

Tutte le formule A e C consentono di ri-ordinare i risultati a seconda delle loro componenti principali presenti nel CBR: queste sono **profili**, **gruppi** e **tag**. Ad esempio, se si chiede il ranking per profili, i risultati vengono ri-ordinati mettendo più in alto quelli con valore di profilo maggiore.

È anche possibile chiedere di vedere i risultati ordinati per default. Questa selezione esclude la componente di CBR, e presenta i risultati facendo delle normali query.

Allo stato attuale, per le formule A1, A2, C1 e C2 vengono presentati solo i primi 200 risultati.

Durante il test sul campo (Edizione di IoLavoro di Ottobre 2013) si sono introdotte molte novità nella definizione dei dati, che ha comportato una serie di modifiche delle formule e del codice corrispondente. Possono essere impostati come campi obbligatori:

- Titolo di studio
- Esperienze formative e qualifiche richieste
- Lingua straniera (ogni singola lingua seleziona può essere definita come obbligatoria)
- Disponibilità a trasferte



- Patente
- Patentino

Queste stesse voci possono essere anche preferenziali.

Inoltre ci sono le categorie protette (invalido, non vedente, tossicodipendente, ...) e lo status lavorativo (disoccupato, in mobilità, ...), che sono diventati elementi di selezione a priori obbligatori se impostati dall'azienda.

Infine, l'esperienza è ora un attributo legato a ogni singolo profilo inserito dell'utente, che può scegliere tra: nessuna, 0-6 mesi, 6 mesi -1 anno, 1 anno-5 anni, 5-10 anni, più di 10 anni.

Dal punto di vista dell'interazione, lo script suggerisce solo lavoratori che non si siano già autocandidati per un annuncio e non suggerisce più alle aziende un candidato se lo ha già contattato. Ad esempio, se un'azienda ha già chiesto un colloquio con un potenziale barman, costui non verrà più suggerito. Idem per i lavoratori. Infine l'utente può chiedere di essere ESCLUSO dai suggerimenti (modalità privata), quindi può soltanto autocandidarsi ma le aziende non vedranno mai il suo profilo tra i lavoratori suggeriti.

Il progetto descritto in questo report è stato implementato nella versione attuale del sito www.lavorointeractive.org (ottobre 2013). Sono stati implementati e testati (sull'edizione di ottobre 2013 di IoLavoro) le principali funzionalità del sistema. In linea di principio, il sistema è stato costruito con criteri di generalità, anche se il modello parametrico e soprattutto il modello valorizzato potrebbero risentire di un problema di *overfitting* sulle situazioni di addestramento di IoLavoro.

In ogni caso, il progetto Lavoro Interactive 3.0 costituisce il punto di partenza per un insieme di possibili sviluppi, grazie al paradigma su cui è costruita l'impalcatura concettuale (il *case-based reasoning*) dell'incontro domanda-offerta di lavoro. Di seguito, si riportano alcuni sviluppi possibili del progetto.

Adattamento agli utenti: Il modello utente sopra descritto fotografa un modello appena inizializzato, dove gli utenti non hanno ancora eseguito nessuna azione sul sistema – oltre alla registrazione. Nel momento in cui l'utente inizia ad interagire con il sistema, se il sistema registrasse le azioni fatte dall'utente, il modello utente potrebbe tenere conto anche delle preferenze implicitamente espresse dagli utenti nei confronti degli annunci e delle aziende, deducibili attraverso la sua interazione con il sistema. Per esempio, attraverso la visualizzazione di un annuncio riferito a un determinato profilo professionale, l'utente implicitamente esprime una preferenza per quel profilo, e dunque per il gruppo e per il settore cui l'annuncio appartiene. L'aggiornamento del modello utente sulla base del comportamento dell'utente permette al sistema di "apprendere" le preferenze dell'utente attraverso la sua osservazione; attraverso apposite "regole di aggiornamento del modello utente", il profilo del candidato (o dell'azienda) evolve man mano che interagisce con il sistema, che ne apprende nel tempo la struttura di preferenze.

Si noti che questo apprendimento sarebbe possibile in quanto sia le istanze degli annunci che quelle delle aziende sono categorizzate in gruppi. Dunque da un'azione su una di queste istanze il sistema potrebbe inferire in maniera implicita, e con peso dipendente dal tipo di azione, un possibile interesse dell'utente verso il gruppo in cui l'istanza è classificata. Le azioni che possono essere tracciate dal sistema sono elencate sotto. I diversi tipi di azioni registrate dal sistema sono pesate in modo diverso a seconda dell'evidenza che forniscono in termini di interessi dell'utente verso un determinato gruppo/settore cui l'azione si riferisce. Ad esempio, se l'utente clicca un annuncio, o su un'azienda, appartenente al settore agroalimentare verrà aggiornato l'interesse dell'utente verso quel settore. Di seguito le azioni e il loro peso corrispondente:

- click su azienda: 0,3
- click su annuncio: 0,3
- inserimento annuncio nel dossier: 0,8
- candidatura annuncio: 0,9

Il modulo di aggiornamento del sistema per ogni azione applicherebbe una variante della seguente formula:

$$computedValue = \frac{\sum_{i=1}^{actions} count(i) \times actionWeight}{actions}$$

Per ciascun tipo di azione i riferita ad un dato profilo/gruppo, $count(i)$ è il numero totale di azioni di effettuate dall'utente e $actionWeight$ è il peso di ogni azione i . La formula somma i risultati del prodotto di $count(i)$ per $actionWeight$, calcolato per ciascuna azione i ; i varia da 1 a $actions$ che è il numero totale di azioni fatte dall'utente su ogni profilo/gruppo. La somma totale viene quindi divisa per $actions$ che rappresenta il numero totale di azioni fatte dall'utente su ogni profilo/gruppo. Così, per ogni profilo/gruppo del modello utente, $computedValue$ rappresenta la probabilità per quel profilo/gruppo, calcolato in base al comportamento dell'utente rispetto a quel gruppo/settore (ossia in base alle sue azioni).

Applicando questa regola a tutti i profili/gruppi presenti nel modello di utente, il modulo di aggiornamento ottiene una nuova distribuzione di probabilità per tutti gli interessi dell'utente. L'aggiornamento potrebbe avvenire ogni 10 azioni dell'utente. La nuova distribuzione di probabilità viene poi combinata con quella già presente nel modello utente per aggiornarlo. Per ogni elemento inferito del modello di utente (profilo/gruppo) e dunque contenente una distribuzione di probabilità, l'aggiornamento avviene come segue:

$$newValue = computedValue * w_1 + currValue * w_2$$

dove $computed value$ rappresenta il valore di confidenza calcolato in base alle ultime azioni dell'utente, mentre $currValue$ è la confidenza già presente nel modello dell'utente e che deve essere aggiornata.

w_1 e w_2 , che devono sommare a 1, e sono usati per dare un peso diverso ai due contributi. All'inizio dell'interazione w_1 è più vicino a 1 ($w_1=0,9$) e w_2 è vicino a 0 ($w_2=0,1$). Con il crescere delle azioni dell'utente nel corso del tempo w_2 viene progressivamente aumentato, mentre w_1 viene diminuito. In particolare ad ogni update $w_1=(w_1-0,1)$ mentre ($w_2= w_2 + 0,1$) fino a quando $w_1=0,1$ e $w_2=0,9$.

Interazione conversazionale con il sistema. Il paradigma del case based reasoning si presta in modo particolare a uno stile di interazione tra utente e sistema più stretto di quello attuale. Ad esempio, sottoponendo le proprie scelte (i suggerimenti) al giudizio dell'utente, il sistema potrebbe sfruttare il feedback dell'utente per adattare il proprio comportamento alle preferenze espresse attraverso il feedback. Secondo questa strategia, utente e sistema collaborano nello "scoprire" progressivamente la struttura di preferenze dell'utente attraverso l'interazione sui singoli casi (annunci o candidati), attraverso un feedback implicito (feedback positivo e negativo sui singoli suggerimenti) o esplicito (eliminazione di feature preferenziali o ri-ordinamento del peso relativo delle feature).

Maggiore utilizzo della semantica. Il sistema attuale si avvale della semantica in due diverse accezioni, cioè nel trattamento dei tag (che vengono confrontati secondo un modello che si basa su un insieme di *feature* linguistiche che includono la sfera semantica di termini di cui il tag è composto) e nell'utilizzo di un ragionamento di tipo tassonomico nella "pesatura" dei profili professionali nella descrizione del candidato. In entrambi i casi, il sistema potrebbe

avvalersi di strumenti più sofisticati. Da un lato, il confronto tra i tag di annuncio e candidato potrebbe avvalersi di “vocabolari” specializzati, da acquisire tramite la disamina di altre iniziative nazionali nell’ambito della ricerca di lavoro; dall’altro, la conoscenza sui profili professionali potrebbe essere aumentata in modo da includere, oltre alle semplici etichette attuali, descrizioni strutturate dei profili stessi, basate su uno schema di competenze, requisiti, ecc. (si vedano ad esempio le iniziative europee EQF e ECVET).

Suggerimento di tag. Il suggerimento di tag si basa attualmente su un approccio di collaborative filtering basato su misure di tipo statistico. Tuttavia, come evidenziato nel presente report, la disponibilità di un insieme più ampio di tag, collezionato durante l’utilizzo del sistema da parte degli utenti, permetterebbe di potenziare questo approccio “scoprendo” correlazioni semantiche tra i tag e la descrizione dei candidati. Tali correlazioni potrebbero quindi essere usate proficuamente ai fini del suggerimento di tag.

Case-based reasoning. L’approccio stesso di Case-based reasoning (CBR) utilizzato potrebbe essere raffinato andando a memorizzare l’esperienza passata degli utenti, ossia andando a memorizzare sia la descrizione del problema che la soluzione applicata in quel contesto. Ad esempio, quando un utente con determinate caratteristiche viene selezionato per un certo annuncio da un’azienda, utenti simili a lui potrebbero essere automaticamente proposti all’azienda. In questo modo il CBR supporterebbe il cosiddetto apprendimento dall’esperienza (learning from experience) usando questi passi fondamentali *retrieve, reuse, adaptation and retain* (recupera, riutilizza, adatta e mantieni)¹⁹. L’adattamento è poi ulteriormente suddiviso in due sotto-passi: *revise*, dove il sistema si adatta ai vincoli del nuovo problema, e *review*, dove il sistema valuta se la soluzione proposta soddisfa l’utente o sono necessari ulteriori aggiustamenti.

¹⁹ Lorenzi F., Ricci F., Case-based recommender systems: a unifying view. In Proceedings of the 2003 international conference on Intelligent Techniques for Web Personalization (ITWP'03), Bamshad Mobasher and Sarabjot Singh Anand (Eds.). Springer-Verlag, Berlin, Heidelberg, pp. 89-113, 2003. DOI=10.1007/11577935_5
http://dx.doi.org/10.1007/11577935_5



Il software che realizza le funzionalità di suggerimento di tag e matchmaking hanno performance accettabile a partire da 1GB della RAM del server.

In fase di sviluppo, si è scelto di adottare tale limite, abbastanza basso rispetto allo stato dell'arte del web hosting, per ottenere una valutazione delle prestazioni del sistema che fosse il più possibile vicina alle peggiori condizioni reali.

Tale limite è stato impostato nel codice attraverso l'istruzione `ini_set("memory_limit", "1024M")` e deve essere modificato a seconda delle caratteristiche effettive del server di produzione, che si auspicano di gran lunga più elevate.

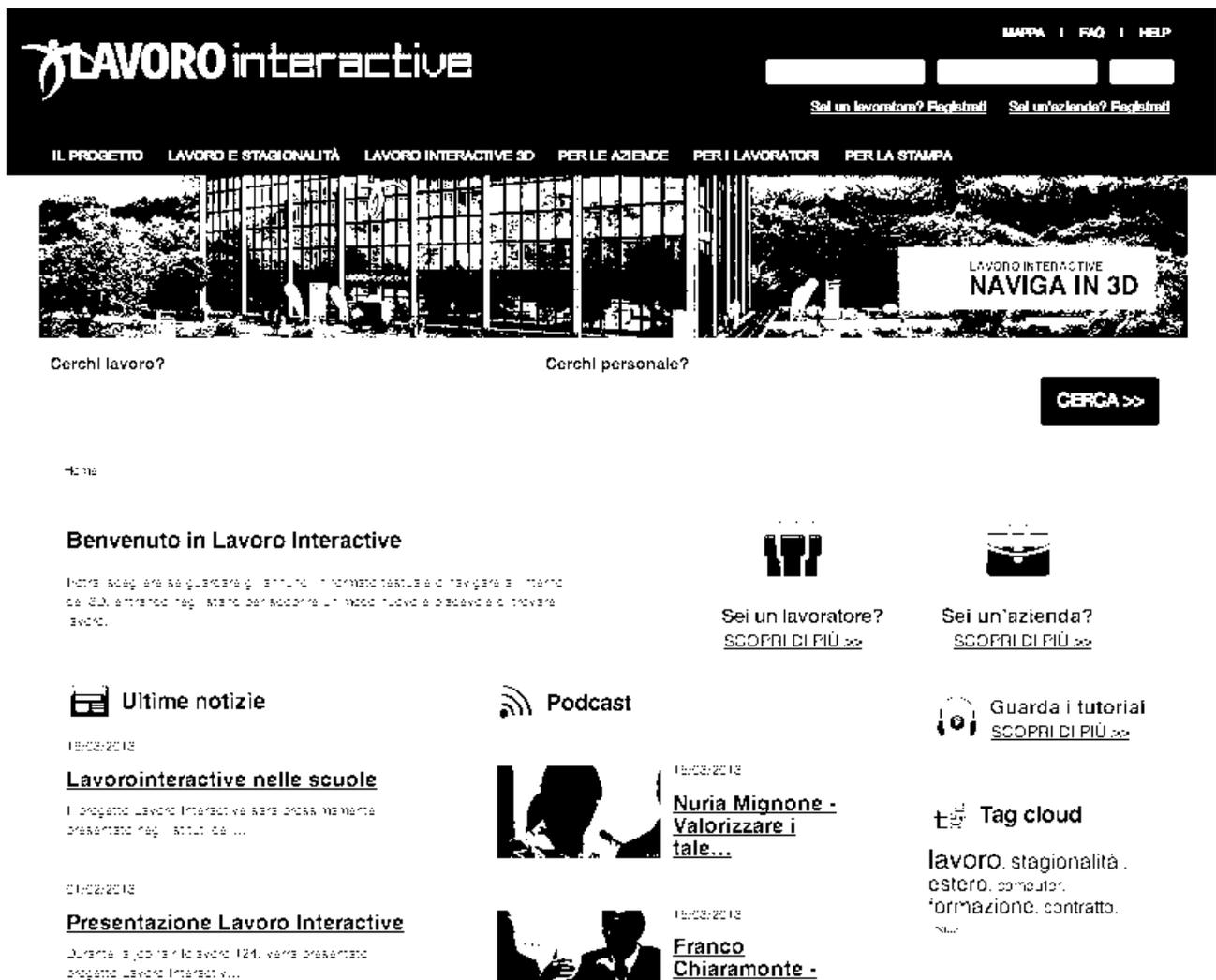
- Ardissono L., Goy A., Tailoring the Interaction with Users in Web Stores. *User Model, User-Adaptive Interaction* 10(4), pp. 251-303, 2000.
- Bentivogli L., Forner P., Magnini B., Pianta E., Revising the wordnet domains hierarchy: semantics, coverage and balancing, In *Proceedings of the Workshop on Multilingual Linguistic Resources (LREC)*, pp. 101--108, ACL, 2004.
- Cattuto C., Benz D., Hotho A., Stumme G.: Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. *International Semantic Web Conference*, pp. 615-631 2008.
- Gatteschi V., Lamberti F., Demartini C., On the Use of Semantic Technologies to Support Education, Mobility and Employability, *Semantic Hyper/Multimedia Adaptation*, pp. 127-150, 2012, Springer.
- Gatteschi V., Lamberti F., Demartini C., van Wezel R., Bettiol S., Exploiting semantics for constructing and comparing occupational and educational-driven qualifications: the TIPTOE project, *Journal of Universal Computer Science*, vol. 18, n.1, 2012.
- Jaeschke R., Marinho L., Hotho A., Schmidt-Thieme L., Stumme G., Tag recommendations in folksonomies, In *Knowledge Discovery in Databases: PKDD 2007*, pp. 506-514, Springer, 2007.
- Koren Y., Bell R. M., Volinsky C., Matrix Factorization. *Techniques for Recommender Systems*. *IEEE Computer* 42(8), pp. 30-37, 2009.
- Lorenzi F., Ricci F., Case-based recommender systems: a unifying view. In *Proceedings of the 2003 international conference on Intelligent Techniques for Web Personalization (ITWP'03)*, Bamshad Mobasher and Sarabjot Singh Anand (Eds.). Springer-Verlag, Berlin, Heidelberg, pp. 89-113, 2003. DOI=10.1007/11577935_5 http://dx.doi.org/10.1007/11577935_5
- Magnini B., Strapparava C., Pezzulo G., Gliozzo A., The Role of Domain Information in Word Sense Disambiguation, *Natural Language Engineering, Special Issue on Word Sense Disambiguation*, 8(4), pp. 359-373, Cambridge University Press, 2002.
- Markines B., Cattuto C., Menczer F., Benz D., Hotho A., Stumme G., Evaluating similarity measures for emergent semantics of social tagging, In *Proceedings of the 18th international conference on World wide web*, pp. 641-650, 2009.
- Navigli R., Word sense disambiguation: A survey, in *ACM Computing Surveys (CSUR)*, vol. 41, n. 2, ACM, 2009.
- Pianta E., Bentivogli L., Girardi C., MultiWordNet: developing an aligned multilingual database, In *Proceedings of the First International Conference on Global WordNet*, Mysore, India, January 21-25, 2002.
- Salton G., McGill M. , *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1984.
- Sigurbjörnsson B., van Zwol R., Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 327-336, 2008. DOI=10.1145/1367497.1367542 <http://doi.acm.org/10.1145/1367497.1367542>.
- Smyth B., Case-Based Recommendation, *The Adaptive Web*, pp. 342-376, 2007.



Smyth B., Briggs P., Coyle M., O'Mahony M. P., A case-based perspective on social web search, In McGinty, L. and Wilson, D. C. (eds.). Case-Based Reasoning Research and Development : 8th International Conference on Case-Based Reasoning, ICCBR 2009, Seattle, USA, July 20-23, Springer, 2009.



In questa appendice presentiamo una serie di immagini (storyboard) che rappresentano lo studio dell'interfaccia, qui illustrate già a seguito dell'intervento dei grafici. Sia l'interfaccia sia l'interazione sono in fase di affinamento e debug all'epoca di scrittura di questa documentazione.



(click su Gestione annunci)

LAVORO interactive MAPPA | FAQ | HELP

Ciao System Srl Sei nella tua MyPage

IL PROGETTO | LAVORO E STAGIONALITÀ | LAVORO INTERACTIVE 3D | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

Benvenuto **SYSTEM Srl**

Il mio profilo | Gestione annunci | La mia cronologia

Ultimi annunci | Nuove autocandidature (20) | Situazione annunci

Hai inserito 15 annunci

- 5 validati e pubblicati
- 2 in attesa di valutazione
- 1 sospeso
- 5 scaduti
- 2 in scadenza

>> Vedi tutti

Numero	Titolo	Scadenza	Stato	Visualizzazioni	Contatti	Altre info
1	88 - Addetto segreteria	06/06/2013	No	15	30	vedi
2	10 - Animatore	14/05/2013	No	22	Calcola	vedi
3	51 - Cuoco	13/05/2013	No	13	23	vedi
4	902 - Cuoco	27/04/2013	No	24	30	vedi
5	76 - Cameriere	26/04/2013	Si	62	29	vedi

(click su selezione di un annuncio)

LAVORO interactive MAPPA | FAQ | HELP

Ciao System Srl Sei nella tua MyPage

IL PROGETTO | LAVORO E STAGIONALITÀ | LAVORO INTERACTIVE 3D | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

Benvenuto **SYSTEM Srl**

Il mio profilo | Gestione annunci | La mia cronologia

Annunci pubblicati | Annunci in attesa di verifica | Annunci scaduti e sospesi

Numero	Titolo	Scadenza	Stato	Visualizzazioni	Contatti	Altre info
1	88 - Addetto segreteria	06/06/2013	No	15	30	vedi
2	10 - Animatore	14/05/2013	No	22	Calcola	vedi
3	51 - Cuoco	13/05/2013	No	13	23	vedi
4	902 - Cuoco	27/04/2013	No	24	30	vedi
5	76 - Cameriere	26/04/2013	Si	62	29	vedi



(click su Autocandidature)

The screenshot shows the 'LAVORO interactive' website interface. At the top, there is a navigation bar with links for 'MAPPA', 'FAQ', and 'HELP'. Below this, a header section includes 'Ciao System Srl' and 'Sei nella tua MyPage'. The main navigation menu contains 'IL PROGETTO', 'LAVORO E STAGIONALITÀ', 'LAVORO INTERACTIVE 3D', 'PER LE AZIENDE', 'PER I LAVORATORI', and 'PER LA STAMPA'. The central banner features a 3D navigation interface with the text 'LAVORO INTERACTIVE NAVIGA IN 3D'. Below the banner, a user profile section for 'Benvenuto SYSTEM Srl' includes icons for 'Il mio profilo', 'Gestione annunci', and 'La mia cronologia'. The main content area displays 'Annuncio n. 88 - Addetto segreteria' with a sub-section for 'Autocandidature' (3) and 'I candidati del sistema'. A 'Chiudi' button is visible in the top right. A detailed description of the job role is provided below the announcement title.

(click su visualizzazione dei candidati selezionati e ordinati dal sistema)

The screenshot shows the 'LAVORO interactive' website interface, specifically the 'Candidati del sistema' section for 'Addetto segreteria'. The layout is similar to the previous screenshot, but the 'Autocandidature' count is 3 and the 'Candidati del sistema' count is 15. A 'Chiudi' button is present. Below the main content, there is a section titled 'Candidature 1-10 su 15 >>'. A table lists the candidates:

Rank	Candidate Name	Score	Date	Status	Action
1	Giovanni Ross	93	06/06/2013	—	Vedi CV
2	Alberto Bianchi	44	14/05/2013		Vedi CV
3	Carlo Giacinto	12	13/05/2013		Vedi CV
4	Francesco Verdi	58	27/04/2013	┌	Vedi CV Aggiungi nota

LAVORO interactive MAPPA | FAQ | HELP

Ciao System Srl
Sei nella tua MyPage

IL PROGETTO | LAVORO E STAGIONALITÀ | LAVORO INTERATTIVE 3D | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

Benvenuto
SYSTEM Srl

Nuovo profilo

Gestione annunci

La mia cronologia

Numero annunci

Annuncio n. 85 - Addetto segreteria Autocandidature 3 I candidati del sistema Chiudi

A3 - STUPISCIMI! - Candidature 1-10

[Ricerca base](#)

Numero	Nome Azienda	Candidature	Scadenza	Stato	Info
1	*****	93		<input type="checkbox"/>	Vedi info
2	*****	14		<input type="checkbox"/>	Vedi info
3	Carlo Gia Ic	12	13/05/2013	<input type="checkbox"/>	Vedi CV Aggiungi nota
4	Francesco Verdi	58	27/04/2013	<input type="checkbox"/>	Vedi CV Aggiungi nota

[A large la ricerca](#)

(click su Cerca annunci)

LAVORO interactive MAPPA | FAQ | HELP

Ciao Mario Rossi
Sei nella tua MyPage

IL PROGETTO | LAVORO E STAGIONALITÀ | LAVORO INTERATTIVE 3D | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

Benvenuto
MARIO ROSSI

Nuovo profilo

Cerca annunci

Candidature e contatti

Impostazioni

La mia cronologia

Lavoro Interactive ti suggerisce

- * **Annuncio n. 88 - Addetto segreteria**
Nome azienda - Inserito il 26/04/2013
- * **Annuncio n. 7 - 3**
Nome azienda - Inserito il 20/04/2013
- * **Annuncio n. 122 - An malore**
Nome azienda - Inserito il 23/04/2013
- * **Annuncio n. 458 - struttore di macchi**
Nome azienda - Inserito il 23/04/2013

[>> Vedi tutti](#)

Richieste di contatto (10 new)

- **Club vacanze** - il 24/04/2013
per Annuncio n. 7 - Animatore baby club
Vedi richiesta
- **Animazione 80** - il 20/04/2013
per Annuncio n. 145 - Animatore di contatto
- **Club ventana** - il 12/04/2013
per Annuncio n. 11 - Animatore di contatto

[>> Vedi tutti](#)

Il tuo status

- Il tuo CV scade il: 15/08/2013
- Candidature effettuate: 43
- Contatti ricevuti: 32
- Annunci salvati: 76
- Aziende salvate: 15

(click su Selezione profilo – C2)

The screenshot shows the LAVORO interactive website interface. At the top, there is a navigation bar with the logo and menu items: IL PROGETTO, LAVORO E STAGIONALITÀ, LAVORO INTERACTIVE 3D, PER LE AZIENDE, PER I LAVORATORI, and PER LA STAMPA. A user profile for Mario Rossi is visible in the top right corner. Below the navigation bar is a large banner image with the text 'LAVORO INTERACTIVE NAVIGA IN 3D'. Underneath the banner are several utility icons: 'Mio profilo', 'Cerca annuncio', 'Candidature e contatti', 'I miei annunci', and 'Cambia immagine'. Below these icons are two tabs: 'Annunci suggeriti' and 'Ricerca libera'. A message states 'Lavoro Interactive ti suggerisce:'. There are three sub-sections: 'SUGGERIMENTI DI BASE (C1) - Annunci 1-10 di 30 >>', 'Selezione per profilo (C2)', and 'Stupiscimi'. The 'SUGGERIMENTI DI BASE (C1)' section contains a table of job suggestions:

Rank	Job Title	Count	Company	Action
1	Addetto segreteria	88	Uripo	Vedi annuncio
2	Animatore	10	Club Med	Vedi annuncio
3	Cuoco	51	Estay	Vedi annuncio
4	Cuoco	902	Club Vacanze	Vedi annuncio

(click su Stupiscimi – C3)

This screenshot shows the same website interface as above, but with the 'Stupiscimi' tab selected. The message now reads 'Lavoro Interactive ti suggerisce:'. The sub-sections are 'Suggerimenti di base' and 'Stupiscimi'. The 'SUGGERIMENTI su PROFILO CUOCO (C2) - Annunci 1-10 di 24 >>' section contains a table of job suggestions:

Rank	Job Title	Count	Company	Action
1	Cuoco	88	Uripo	Vedi annuncio
2	Cuoco	10	Club Med	Vedi annuncio
3	Cameriere	51	Estay	Vedi annuncio
4	Chef	902	Club Vacanze	Vedi annuncio

MAPPA | FAQ | HELP

Ciao Mario Rossi

Sai nella MyPage

IL PROGETTO | LAVORO E STAGIONALITÀ | LAVORO INTERACTIVE 3D | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

LAVORO INTERACTIVE
NAVIGA IN 3D

Benvenuto
MARIO ROSSI

Meo profilo

Cerca annuncio

Candidature e contatti

Le mie aziende

La mia cronologia

Annunci suggeriti Ricerca libera

Lavoro Interactive ti stupisce!

Suggerimenti di base Seleziona per profilo/tendenza

C3 - STUPISCIMI! - Annunci 1-10

1	2	3	4
Addetto segreteria	88	Uripo	Vedi annuncio
Animatore	10	Club Med	Vedi annuncio
Cuoco	51	Eata y	Vedi annuncio
Cuoco	902	Club Vacanze	Vedi annuncio

MAPPA | FAQ | HELP

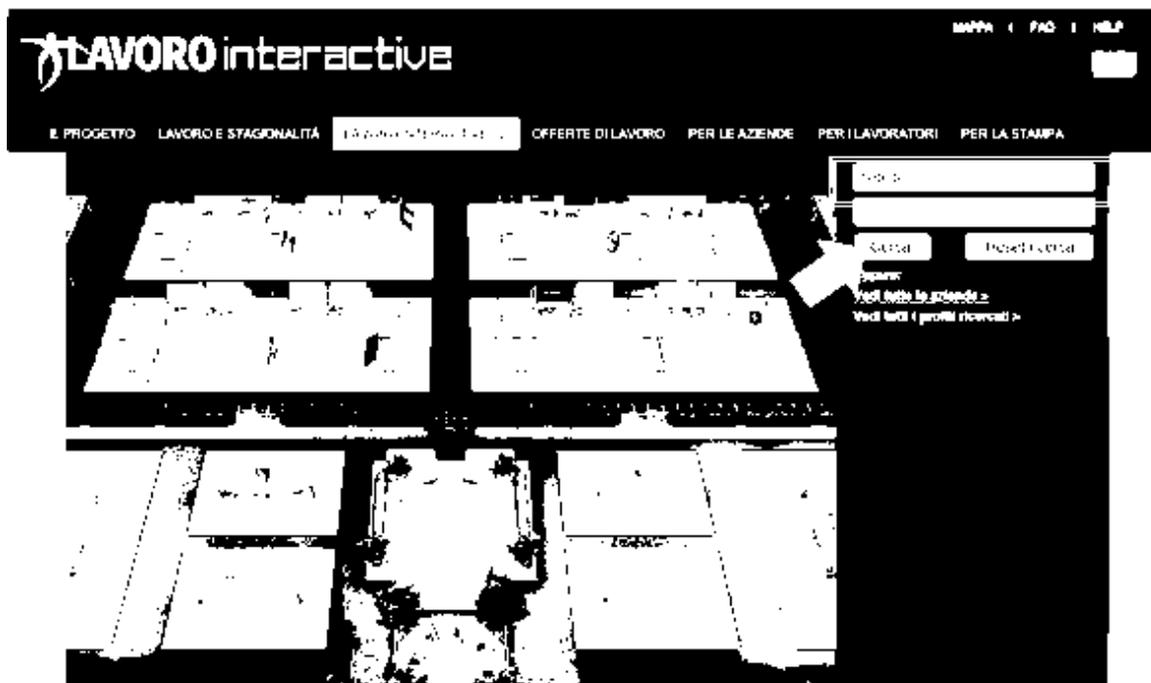
Ciao

IL PROGETTO | LAVORO E STAGIONALITÀ | **LAVORO INTERACTIVE 3D** | OFFERTE DI LAVORO | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

Cerca Reset ricerca

Oppure:
[Vedi tutte le aziende >](#)
[Vedi tutti i profili ricercati >](#)

(click per ricerca libera "cuoco")





LAVORO interactive MAPPA | FAQ | HELP

Ciao Marco,
[Va alla MyPage >](#)

IL PROGETTO | LAVORO E STAGIONALITÀ | **USI E ALTERNANZE** | OFFERTE DI LAVORO | PER LE AZIENDE | PER I LAVORATORI | PER LA STAMPA

SUGGERIMENTI PIÙ PER TE

Ciao Marco,
ecco gli incarichi suggeriti per te:

PIZZAIO	
Old Man	<input type="text"/>
ARMATORE	
Fall Time	<input type="text"/>
RECUPEROIST	
Your Friend	<input type="text"/>
BAGNAIO	
Yellow	<input type="text"/>
CAMBIERE	
Edley	<input type="text"/>
COMMISSO	
Edley	<input type="text"/>