

PBS Talent WORDS (v11): algoritmi proprietari di Induzione Semantica per l'analisi delle candidature (profili e documenti)



marzo 2022



Agenda

- PBS in breve
- **PBS Talent Words**
 - Concetti
 - Architettura
 - Demo e istruzioni per l'uso
 - Licensing



Primo Bonacina, Founder & CEO of PBS



PBS: 4 Business Units

PBS Talent (Digital Recruiting, AI/HR Software, People Assessment)

Podcasting (RADIO IT/Podcast Italia Network platforms)

Online Advertising (Google Ads)

Tailored Consulting (Digital, Events, Keynote, Management, LinkedIn, Workshops, Languages, Custom Projects ...)



PBS: business highlights

2014

founded

1300

transactions

100%

service revenue

1.2M

web site top rank

100+

podcasting sponsors

370

customers

550

job posts
(digital recruiting)

€2.3M

total revenue

€458k

12-month revenue

100+

Business events

38

managed brands (Google Ads)

17

managed companies
(social media)

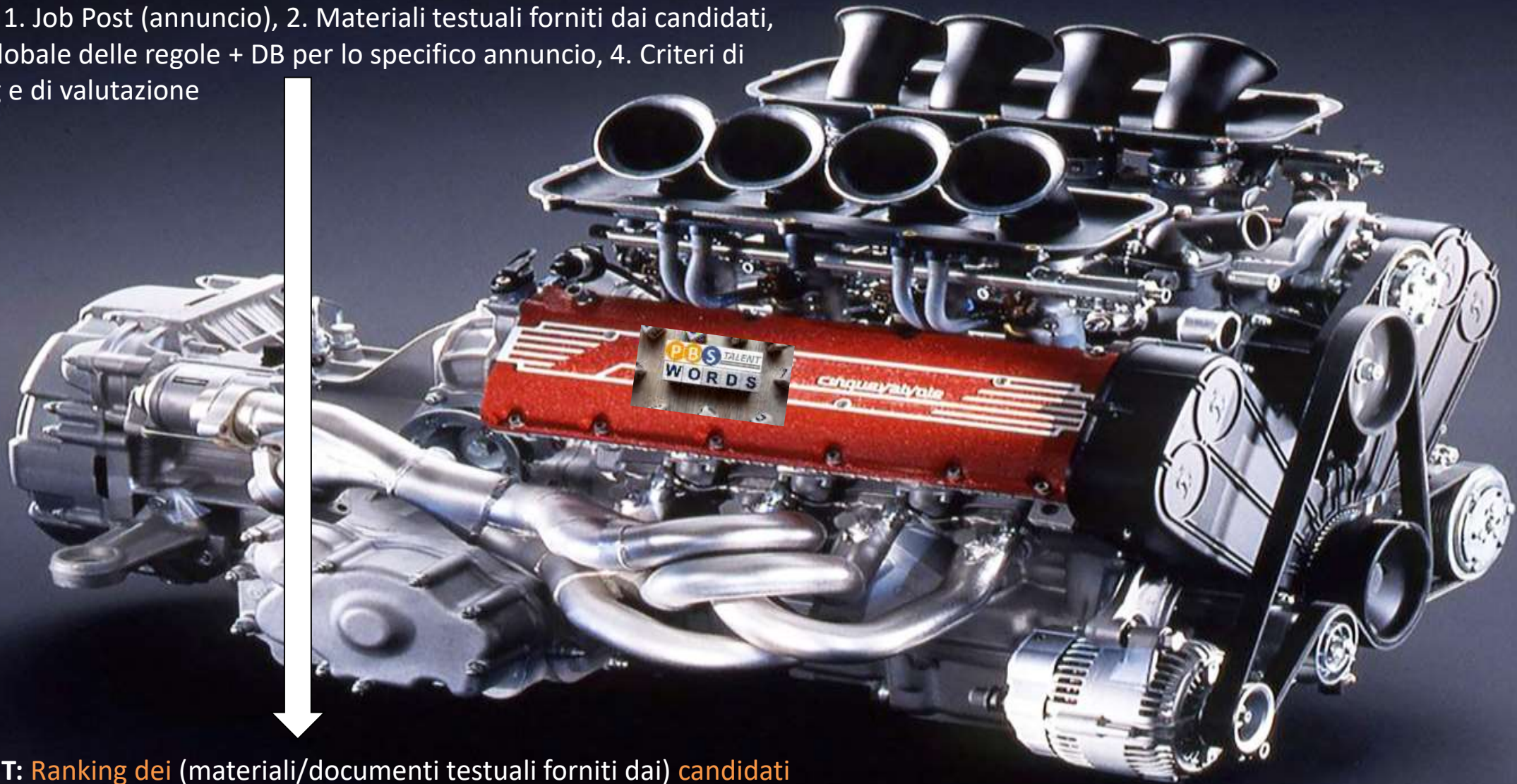
20

countries of revenue

Cosa è PBS Talent Words?

Una macchina di **ranking semantico** di documenti **testuali**

INPUT: 1. Job Post (annuncio), 2. Materiali testuali forniti dai candidati,
3. DB globale delle regole + DB per lo specifico annuncio, 4. Criteri di
ranking e di valutazione



OUTPUT: **Ranking dei** (materiali/documenti testuali forniti dai) **candidati**

A cosa serve

1. Selezione su *estrazioni database* e, in generale, su *quantitativi massivi*: filtrare profili sopra un certo punteggio
2. **Supporto alla decisione manageriale**: passaggio dalla long list alla short list, presentazione a clienti e stakeholder
3. **Analisi di documenti testuali**: anche fuori dal recruiting



A chi è primariamente indirizzato

- Chi si occupa di recruiting in azienda (HR, management)
- Agenzie di ricerca e selezione

Target Audience



L'intuizione alla base: l'induzione secondo Aristotele

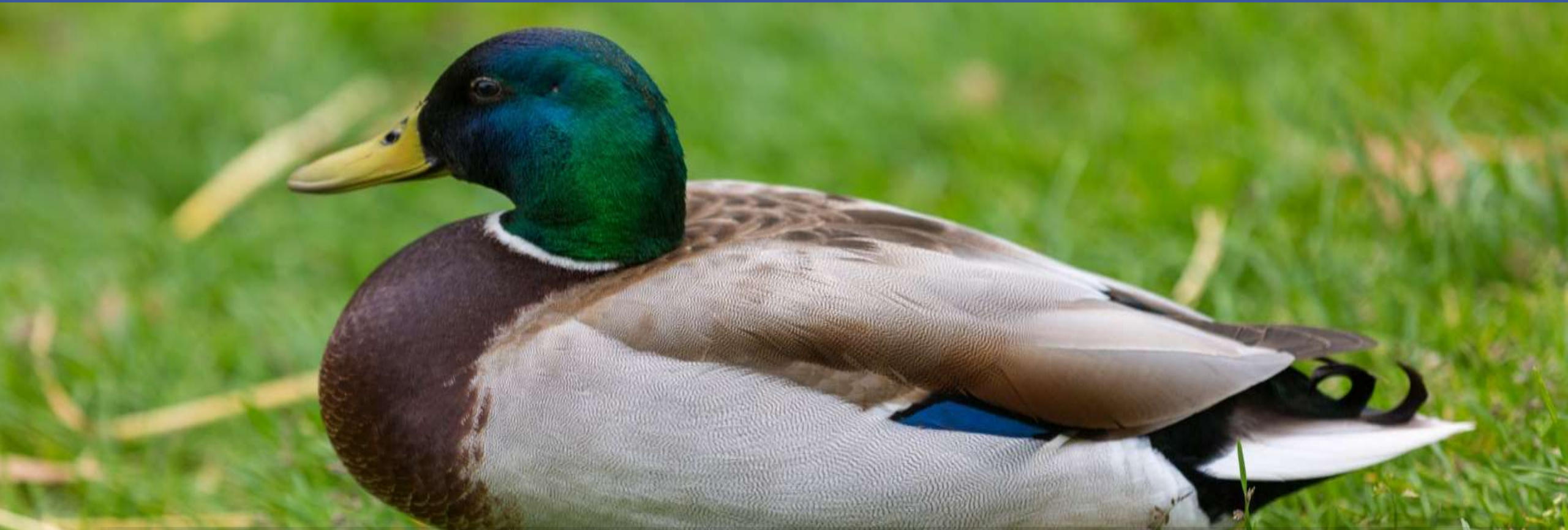
L'induzione in Aristotele [modifica | modifica wikitesto]

Uno dei primi filosofi a ricorrere a questo concetto fu **Aristotele**, il quale, attribuendo a **Socrate** il merito di averla scoperta,^[1] sosteneva che l'induzione fosse, appunto, «il procedimento che dai particolari porta all'universale».^[2]

Secondo Aristotele, la **conoscenza** umana si può svolgere in due direzioni, scegliendo una delle due seguenti strade: avere una prima conoscenza **sensibile** del particolare e da questa risalire all'universale (via dell'induzione, appunto), o seguire la strada opposta, cioè partire dall'universale per andare al particolare (via della **deduzione**).



Ancora meglio: **Test dell'anatra** (ragionamento abduttivo)



Con **test dell'anatra** o **duck test** si indica in modo umoristico una forma di **ragionamento abduttivo**. La proposizione più comune attraverso cui è espresso è la seguente:

(EN)

«If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck.»



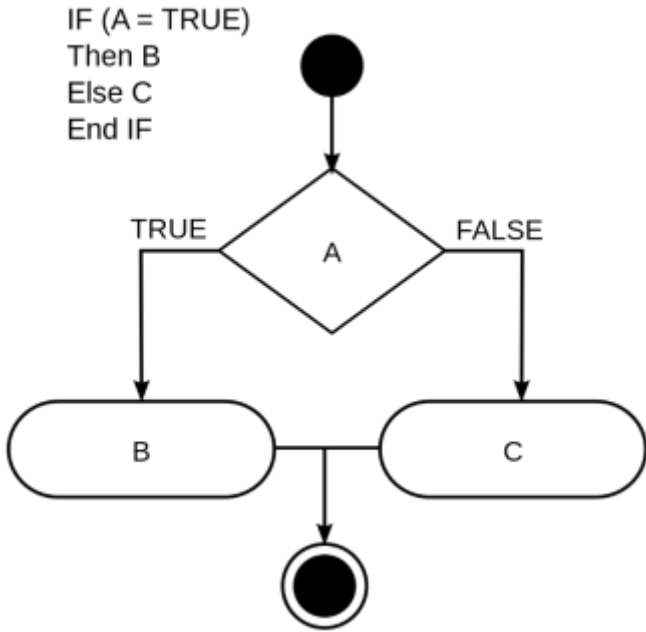
«Se sembra un'anatra, nuota come un'anatra e starnazza come un'anatra, allora probabilmente è un'anatra.»



Conseguenza logica del test è che una persona può identificare un soggetto sconosciuto osservandone le caratteristiche abituali. È a volte utilizzato per controargomentare considerazioni ermetiche ed astruse che vorrebbero dimostrare che qualcosa non è ciò che appare.

Problema #1: i metodi di **selezione per keyword** sono rozzi, richiedono di pensare a troppe varianti

```
IF (A = TRUE)  
Then B  
Else C  
End IF
```



- Troppi **sinonimi** per ruoli simili (**Italiano e Inglese**): Developer, Sviluppatore, , Programmer, Programmatore, Analista, Analyst, Architect ...
- Non serve una selezione, serve un **ranking** che supporti le persone (**e poi le persone decidono**)





Problema #2: i metodi **generalisti** mancano di sufficienti input specifici, danno riscontro approssimativo

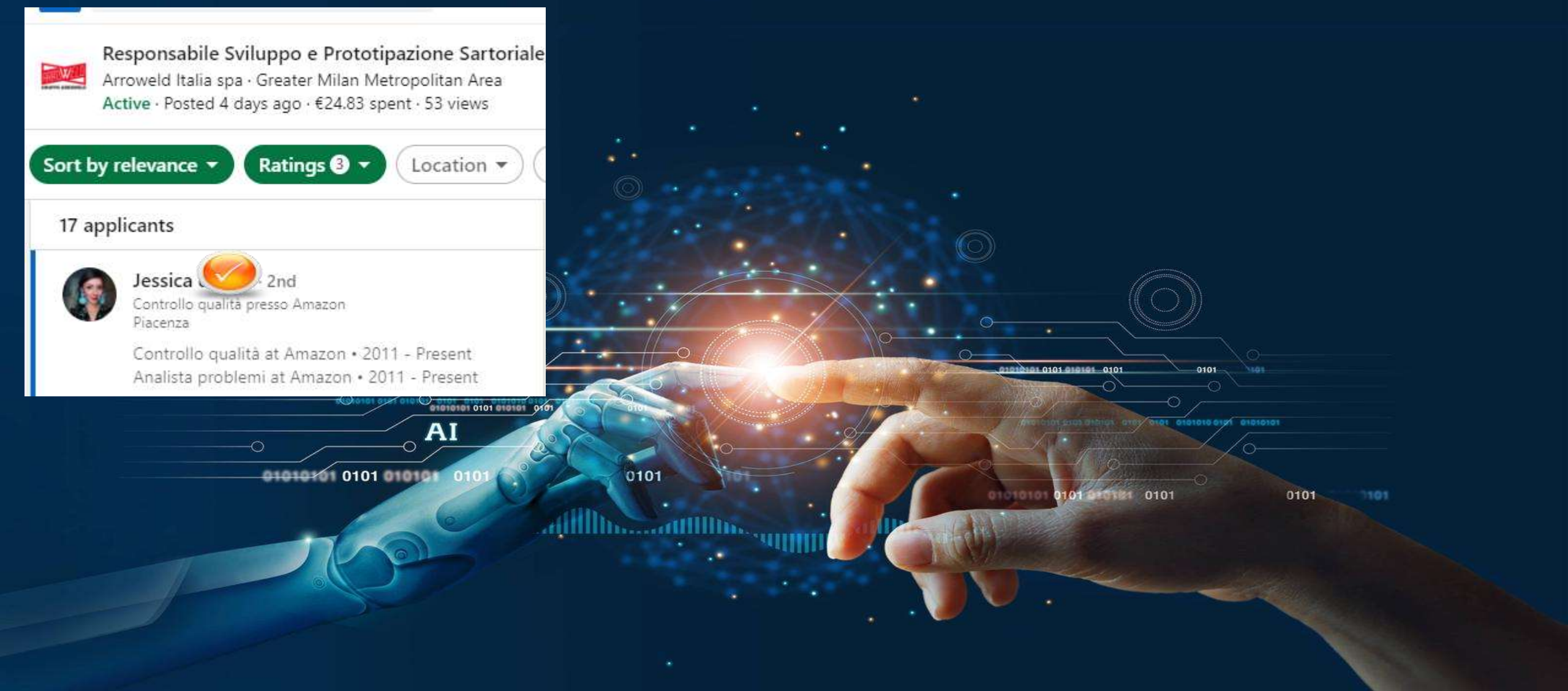
Responsabile Sviluppo e Prototipazione Sartoriale
Arroweld Italia spa · Greater Milan Metropolitan Area
Active · Posted 4 days ago · €24.83 spent · 53 views

Sort by relevance ▾ Ratings 3 ▾ Location ▾

17 applicants

 Jessica  2nd
Controllo qualità presso Amazon
Piacenza

Controllo qualità at Amazon · 2011 - Present
Analista problemi at Amazon · 2011 - Present



10 motivi per un sistema ad hoc



1. Usato **OGNI** giorno; **Alla base di PBS Talent**
2. L'intelligenza è **nell'input**, per il resto è un parser
3. **Super rapido**: pochi secondi, pochi tocchi di tastiera, veloce e automatizzato (**max 2 minuti per sessione**)
4. Chiaro, **semplice**, limitato, ben definito
5. **Disponibilità sorgenti** per crescere ed evolvere
6. Interfaccia utente in **Italiano**
7. **Analizza testi di qualunque tipo** (CV, profilo LinkedIn in PDF, interviste, materiale aggiuntivo): un candidato può venire **descritto da più documenti**
8. **Database regole multilingua** (standard: Italiano e Inglese; Altre lingue si aggiungono editando le regole); **Sistema indipendente dalla lingua**
9. Il database delle regole incrementa: **intelligenza e completezza aumentano a ogni incarico**
10. **Versatile**: le regole possono descrivere **qualsiasi ruolo/mercato/settore**



Un po' di storia

- Nasce ad aprile 2020 come modulo di PBS Talent
- Ridisegno a gennaio 2021
- v11 a marzo 2022
- Testato in centinaia di incarichi



Le attività PBS Talent Words si innestano naturalmente nel flusso di lavoro del recruiter

1. Scrittura annuncio (ricerca profilo) e sua pubblicazione
2. Creazione database di regole per lo specifico annuncio, a partire da un database globale
3. Raccolta documenti dai candidati (da database dei candidati, annunci ...)
4. Ranking dei candidati
5. Ottimizzazione database globale regole

... e si riparte, ancora più informati e ottimizzati



AI e Semantica: un esempio di lavoro

Sr. Key Sales Account Managers, Enterprise Cybersecurity & Data Center Infrastructure (Roma)

Core Sistemi nasce nel 2005, specializzandosi nelle soluzioni infrastrutturali IT per grandi aziende e pubbliche amministrazioni, guidate nella loro trasformazione digitale tramite la proposizione di soluzioni Enterprise IT all'avanguardia. Dotato di uno staff tecnico specializzato, Core Sistemi ha ottenuto risultati di successo con alcune delle principali aziende italiane, collaborando in progetti tecnologicamente significativi alla continua ricerca dell'innovazione. Core Sistemi si guadagna la fiducia dei clienti garantendo un elevato livello di servizio e un chiaro ritorno sull'investimento. Il business di Core Sistemi è incentrato sulle linee Cybersecurity, Data Center e Managed Services & Solutions. Core Sistemi, certificata ISO 27001/9001, impiega 35 dipendenti per un fatturato di €6M+ in fortissima crescita

All'interno di un progetto di espansione, Core Sistemi ricerca alcuni Sr. Key Sales Account Managers per la vendita di progetti di Enterprise Cybersecurity & Data Center Infrastructure su aziende di taglio Enterprise pubbliche e private primariamente nel Lazio e regioni limitrofe

Le figure riporteranno al CEO. La maggioranza dell'attività verrà svolta presso i clienti, recandosi occasionalmente a Roma Tor di Valle (Cenci/Spinaceto) quando necessario

Compiti e Responsabilità

- Responsabilità di un proprio budget di vendita e margine
- Conoscenza del portafoglio di soluzioni infrastrutturali con capacità di presentare il progetto e intrattenere un dialogo approfondito con le figure chiave tecniche (CISO/CIO/IT Purchasing) e non del cliente
- Ricerca, ingaggio, sviluppo e presidio in autonomia di un proprio portafoglio clienti: aziende di taglio Enterprise pubbliche e private (1.000+ posti di lavoro, €200M+ fatturato)
- Approccio consulenziale lungo il ciclo di vendita con gestione di offerte, in affiancamento alla prevendita e ai vendor
- Conduzione della vendita: gestione portafoglio clienti, pianificazione visite, presentazione soluzioni, generazione pipeline, mappatura/qualificazione opportunità, forecasting/CRM, produzione offerte, negoziazione ordine
- Responsabilità e gestione delle problematiche del cliente (e delle relative opportunità) nella sua totalità
- Acquisizione delle principali certificazioni commerciali dei vendor
- Cooperazione con il team di delivery per l'implementazione delle soluzioni
- Interazione con le altre funzioni aziendali in logica di team working

Competenze e Requisiti

- Conoscenza, relazione e presidio di un adeguato portafoglio clienti sul territorio
- Esperienza nella vendita di progetti infrastrutturali IT su grandi aziende pubbliche e/o private
- Esperienza su alcune delle seguenti aree tecnologiche: Cybersecurity, Networking, Cloud, IT Infrastructure, Data Center, Managed Services
- Esperienza su alcuni dei seguenti vendor (o comparabili): Fortinet, Check Point, CyberArk, Dell Technologies, Fore Scout, Cisco, HPE, Gemalto, Vectra AI, Proofpoint, Nutanix; La presenza di certificazioni costituirà un plus
- Conoscenza della lingua inglese

Caratteristiche vincenti

- Questo è un ruolo "100% Enterprise IT sales" che richiede basi tecniche ma soprattutto vera passione per la vendita, dedicando la maggioranza del tempo al presidio dei clienti per lo sviluppo di opportunità
- Volontà di eccellere in ambienti dinamici e in mercati in evoluzione, accrescendo continuamente le proprie competenze
- Capacità di far emergere le esigenze latenti del cliente e proporre soluzioni innovative, offrendo un vero servizio e generando una relazione di valore e durata
- Capacità relazionali/empatia, approccio orientato al risultato, capacità di lavorare per obiettivi, serietà/puntualità, energia, standing adeguato
- Capacità di pianificare attività a medio/lungo termine per raggiungere costantemente i propri obiettivi
- Capacità di negoziazione, forecasting, controllo del business
- Disponibilità a continue brevi trasferte in zona

La ricerca è per assunzione immediata a tempo indeterminato (Quadro)

Il pacchetto (fisso + variabile + assicurazione sanitaria + car allowance) potrà soddisfare candidature TOP

1. Annuncio

Parole chiave	Exact_Match
account	yes
account manage	
acqui	
purchas	
amm	
approfondi	
assicur	
insur	
aziend	
bu	
budget	yes

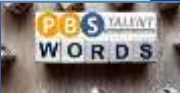
2. Incoro
 keyword annuncio
 («token») e
 database
 regole (3.000)

4. Ranking
 semantico pesato
 (+ criteri
 oggettivi)

Keyword	Weight	Exact_Match	Progr.
account	10	yes	1%
adinet	10		2%
appalt	10		3%
check point	10		4%
cio	10	yes	5%
ciso	10		6%
consys	10		7%
de sistemi	10		8%
erat	10		9%
clber	10		10%
enterprise	10		11%
fore Scout	10		12%
fortinet	10		13%
gar	10		14%
key	10		15%
mepa	10		16%
pa	10	yes	17%
pac	10	yes	18%
pal	10	yes	19%
proofpoint	10		20%
sales	10		21%
secur	10		22%
sicur	10		23%
tele consys	10		24%
welcom	10		25%
cisco	9		25%
commerci	9		26%
difesa	9		27%
firewall	9		28%
Nutanix	9		29%
poste	9		30%
public	9		31%
public	9		32%
anit	9		32%
an	9		33%
asystel	8		34%
cloud	8		35%
darktrace	8		36%
data cent	8		36%
defen	8		37%

3. Messa a punto
 (150+) regole
 semantiche
 per job post

File	Punti	Punti (no ripetizioni)	Numero parole	Numero caratteri
f	578	160	865	5555
f	578	160	865	5555
f	578	160	865	5555
f	578	160	865	5555
f	580	233	645	4882
05.pdf	271	166	295	2041
07.pdf	244	145	302	2059
08.pdf	360	133	467	703
i01.docx	497	21	14	395
i02.pdf	319	21	16	395
i03.pdf	595	247	937	607
i04.pdf	1257	33	3	11697
i05.docx	5	250	73	4846
i06.pdf	593	250	73	4790
i07.pdf	595	262	573	3721
i08.pdf	164	114	427	4348



Long List (decine di profili per job post): Ranking candidati
«normalizzato»: Punti (0-100), Fasce 1-2-3 (esempio: 100-80-60)



Il ranking funziona alla prova dei fatti! (% candidati **incontrati** da clienti PBS Talent)

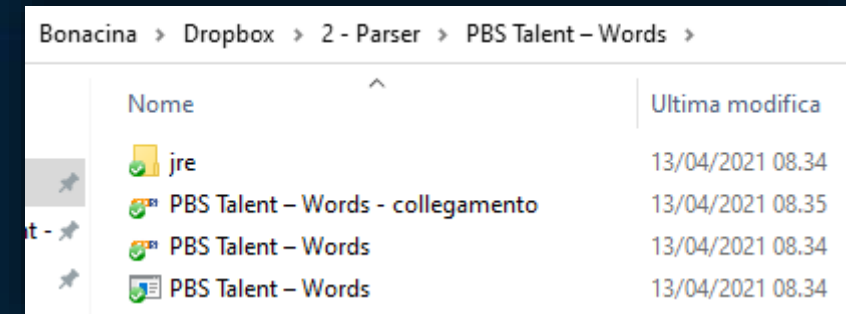


Tour guidato (con schermate e manuale d'uso)



Architettura di sistema

- **Applicazione autosufficiente per Windows;** Tutti i moduli in una directory locale
- Può prendere dati da qualsiasi **directory** (locale o cloud)
 - Con dati locali, non richiede accessi in cloud
 - Elabora e produce output in locale
- Duplicabile su più postazioni
 - NO autenticazione, registrazione, credenziali
- **Non registra dati, non gestisce database (GDPR compliant)**
- **Formati standard**
 - **Input: Microsoft Office (XLS/XLSX, DOC/DOCX), ODT (OpenDocument: OpenOffice/LibreOffice), PDF, bitmap (OCR)**
 - **Output: XLSX (massima manipolabilità)**
- **Semplice, veloce, efficace**
 - Non va in crash, anche a fronte di dati di input non conformi
 - I dati nelle schermate sono ordinabili secondo il criterio preferito
 - Ricorda le ultime impostazioni



Nome	Ultima modifica
jre	13/04/2021 08.34
PBS Talent - Words - collegamento	13/04/2021 08.35
PBS Talent - Words	13/04/2021 08.34
PBS Talent - Words	13/04/2021 08.34

Specifiche tecniche

- Applicazione in Java
- Integrated Development Environment (IDE): Eclipse
- Estrazione dati dai file: libreria Apache Tika
- Estrazione OCR: Tesseract OCR engine
- Riconoscimento lingua: LanguageHandler/OptimizeLangDetector (in Apache Tika)
- Ricerca/conteggio parole chiave: RegEx (Regular Expression) su dati estratti da Apache Tika/Tesseract
- GUI: libreria grafica Java (Swing), plugin WindowBuilder
- Distribuzione: eseguibile tramite Cross-platform Java executable wrapper Launch4j (OpenJDK 11)



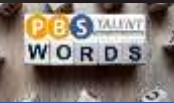
Definizioni semantiche

- Una **parola** è una sequenza di **lettere, numeri e il segno «#»** separate da qualsiasi spaziatura o punteggiatura
- Una **stringa** è una sequenza di parole, usata come **prefisso** (salvo che non sia richiesta in «**Exact Match**»)
 - Esempio: “account manage” è una stringa
 - Nel testo si potranno trovare le sequenze (**prefissi**) “account manager” e “account management”. Salvo che sia impostato l’Exact Match (vedi oltre), entrambe vengono trovate
- Una **regola** può essere una parola o una stringa
 - **il sistema dà un punteggio alle regole riscontrate nei documenti**



Un sistema in 2 semplici comandi

1. **Analisi annuncio:** partendo dal database delle regole, estrae le stringhe contenute nell'annuncio per permettere di costruire un set di regole per lo specifico annuncio
 2. **Punteggio:** sulla base delle regole dell'annuncio, assegna un ranking al candidato (ai documenti forniti)
- **(+ funzione di servizio) Setting:** repository regole/annunci/documenti e soglia per OCR



(funzione di servizio) Scelta directory e setting OCR

- **Documenti:** file forniti dai candidati per ranking
- **Annunci:** job post
- **Parole chiave:** database delle regole (globale e per specifico annuncio)
- **Output:** dove salvare le analisi
- **Setting:** Livello di soglia per OCR (i documenti illeggibili o che non superano questo punteggio vengono analizzati con OCR)

Default modificabile



Analisi Annuncio	Punteggio	Setting
Documenti:	C:\Users\Bonacina\Google Drive\Drive - PBS Talent	
Annunci:	C:\Users\Bonacina\Dropbox\0 - Word\Word - Talent - Annunci	
Parole chiave:	C:\Users\Bonacina\Dropbox\0 - Excel\Excel - Words	
Output:	C:\Users\Bonacina\Downloads	
Puntì OCR:	30	

Database globale delle regole (XLS)

- Ad oggi quasi 3.000 regole o varianti
- Formato semplicissimo (5 colonne XLS)
- Maiuscolo e minuscolo sono irrilevanti (in tutto il progetto)
- Duplicabile/editabile per adattarlo a diversi mercati

Formato:

- **1° colonna («0»)** per facilitare ordine alfabetico oppure «keyword»: lista delle keyword/regole
- **2° colonna («Exact_Match»)**: indica se la regola deve venire riscontrata **solo esattamente oppure anche solo come «prefisso» (inizio parola)**
 - «pa» con Exact_Match trova solo «pa»
 - «pa» senza Exact_Match trova «pa» e «pane», **NON** trova «apatia» (**non è «prefisso»**)
- **3° colonna («Additional»)**: nel caso trovi la 1° colonna, propone anche questa come regola aggiuntiva; Obiettivo: prevedere varianti linguistiche e alternative di scrittura
 - Un'alternativa di «accountant» o «accounting» è «contabil» (variante italiano/inglese)
 - Si possono creare varianti per ogni lingua o contesto
- **4° colonna («Additional #2»)**: nel caso trovi la 1° colonna propone anche questa come regola aggiuntiva
- **5° colonna («Punteggio Base»)**: propone questo punteggio come default al momento di personalizzare le regole
- 6° colonna e seguenti: di servizio e liberamente modificabili
 - 6° colonna: conta le regole
 - 7° colonna: verifica eventuali duplicati
 - Altre colonne: conteggi e statistiche

0	Exact_Match	Additional	Additional #2	Punteggio base		Duplicate?	0	Total
68				10	1	0%	0	
365				5	2	0%	0	
365				5	3	0%	0	
730				5	4	0%	0	
740				5	5	0%	0	
770				5	6	0%	0	
802				5	7	0%	0	
900				5	8	0%	0	
2d				3	9	0%	0	
3cx				5	10	0%	0	
3d				8	11	1%	0	
3ntr				10	12	1%	0	
a10				8	13	1%	0	
a4				5	14	1%	0	
abb	yes			5	15	1%	0	
abbiategrasso				3	16	1%	0	
abbigliamento		apparel	vesti	5	17	1%	0	
abilitazion				3	18	1%	0	
abrasiv				8	19	1%	0	
academ		accademi	universit	5	20	1%	0	
access				4	21	1%	0	
access point				5	22	1%	0	
accessor		tool		3	23	1%	0	
acciai		steel	metal	5	24	1%	0	
accord		agreem		5	25	1%	0	
account	yes			10	26	1%	0	
accountant		contabil		5	27	1%	0	
accounting		contabil		5	28	1%	0	
acer		asus	lenovo	10	29	1%	0	
acet				5	30	1%	0	
achiev		risultat		5	31	1%	0	
achieve				2	32	1%	0	



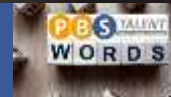
Analisi dell'annuncio

1. Scelta del testo dell'annuncio (DOCX)
2. Scelta del database globale delle regole

In pochi istanti, produzione di un file XLS (directory standard Windows: DOWNLOAD) con la bozza delle regole dello specifico annuncio

Nome	Modificato	Dimensioni	Tipo
2104 Words - LAN & WAN - Sales - Padova	29/04/2021 17.49	34,4 KB	Foglio di lav...
2104 Words - vuoto	29/04/2021 17.46	30,2 KB	Foglio di lav...
2104 Words - LAN & WAN - Networking - Padova	29/04/2021 17.46	32,3 KB	Foglio di lav...
2104 Words - Euris - Atlassian	29/04/2021 17.45	32,4 KB	Foglio di lav...
2104 Words - ICONA - PHP Developers	29/04/2021 17.45	32,5 KB	Foglio di lav...
Words - Keywords	28/04/2021 11.14	97,9 KB	Foglio di lav...
Words - Ricerche	28/04/2021 11.13		Cartella di file

Nome
Analisi annuncio - 30_04_2021-09_49_19



Bozza regole annuncio

- Il file XLS prodotto contiene tutte le regole riscontrate nell'annuncio
- La **1° colonna** indica la **regola**, la **2°** se è stata riscontrata in «**Exact Match**», la **3°** propone il punteggio base (di default), poi liberamente modificabile
- Nel caso trovi una regola in e questa possieda una regola **ADDITIONAL** e/o **ADDITIONAL #2**, la propone subito dopo nel file XLS di output

Parole chiave	Exact_Match	Punteggio base
acqui		2
purchas		2
aditinet		10
consys		10
lunit		10
anal		2
app		1
software		1
sw		1
applica		5
application delivery		20
approfondi		2
archite		8
area		2
arista		5
aziend		2
compan		2
bi		8
business		2
capacit		2
capitolat		2
bid		2
center		3
centr		2
certif		6
cicl		2
cisco	yes	6
client		2
cloud		7
collabora		2
collaud		3
comm		3
commerc		10
commess		8
competen		3
skill		3



Messa a punto (manuale) delle regole, #1

File XLS vuoto con un formato prefissato per ospitare le regole dello specifico annuncio:

1. «0» (per facilitare ordine alfabetico) oppure «**keyword**»: contiene le keyword/regole
2. «**Weight**»: contiene il peso arbitrario che daremo alla regola
3. «**Exact_Match**»: indica se la regola deve venire riscontrata esattamente oppure anche come prefisso
4. 4° colonna e seguenti: di servizio e liberamente modificabili

0	Weight	Exact_Match	Progr.	Duplicate?
			#DIV/0!	1 0
			#DIV/0!	2 0
			#DIV/0!	3 0
			#DIV/0!	4 0
			#DIV/0!	5 0
			#DIV/0!	6 0
			#DIV/0!	7 0
			#DIV/0!	8 0
			#DIV/0!	9 0
			#DIV/0!	10 0



Messa a punto (manuale) delle regole, #2

- Il file XLS prodotto contiene tutte le regole riscontrate poi da mettere a punto manualmente
- Il peso (**Weight**) è arbitrario e varia incarico per incarico (**convenzione: 10 = peso alto; 1 = peso minimo; valori sopra 10: eccezioni e «must»**)
- Il campo **Exact_Match** viene messo a punto parola per parola
- Dallo studio (da parte dell'azienda) dell'annuncio, **emergeranno nuove regole** non ancora presenti nel database globale delle regole; Nel file XLS delle regole dell'annuncio si inseriranno queste nuove regole, poi **da riportare nel database globale delle regole (arricchimento della «knowledge base»)**

	Weight	Exact_Match	Progr.	Duplicate?		
anal	3		1%	1	0	3
applica	4		2%	2	0	7
assist	6		3%	3	0	13
aziend	2		4%	4	0	15
brescia	5		5%	5	0	20
bs	5		6%	6	0	25
client	2		7%	7	0	27
competen	2		8%	8	0	29
css	10		10%	9	0	39
customer	2		11%	10	0	41
degree	3		11%	11	0	44
develop	10		14%	12	0	54
digital	3		15%	13	0	57
erp	5		16%	14	0	62
esper	2		17%	15	0	64
evolu	2		17%	16	0	66
expert	2		18%	17	0	68
fabbric	5		19%	18	0	73
factor	5		20%	19	0	78
formazion	3		21%	20	0	81
front	10		24%	21	0	91
function	3		24%	22	0	94
funzion	3		25%	23	0	97
gestion	3		26%	24	0	100
grafic	3		27%	25	0	103
html	10		29%	26	0	113
improv	4		30%	27	0	117
industr	5		32%	28	0	122
informa	2		32%	29	0	124
innova	2		33%	30	0	126
install	6		34%	31	0	132
interfac	10		37%	32	0	142
java	4		38%	33	0	146
javascript	10		40%	34	0	156
jq	10		43%	35	0	166
jr	8		45%	36	0	174
laure	3		46%	37	0	177
maint	5		47%	38	0	182
manif	5		48%	39	0	187
manufa	5		50%	40	0	192
manuten	5		51%	41	0	197
mes	10	yes	54%	42	0	207



Collocazione in una directory dei documenti da analizzare

- Estrazione da database CV
- Scaricamento manuale profili da LinkedIn
- Documenti forniti dai candidati
- Altre fonti

Nome	Modificato	Dimensioni	Tipo
05	29/04/2021 17.38	63,3 KB	Chrome HT...
i05	29/04/2021 17.38	184 KB	Chrome HT...
04	29/04/2021 09.26	258 KB	Chrome HT...
i04	29/04/2021 09.26	1,00 MB	Chrome HT...
03	28/04/2021 10.31	332 KB	Chrome HT...
i03	28/04/2021 10.31	65,1 KB	Documento ...
02	28/04/2021 10.27	218 KB	Chrome HT...
i02	28/04/2021 10.27	257 KB	Chrome HT...
01	28/04/2021 10.23	265 KB	Chrome HT...
i01	28/04/2021 10.23	167 KB	Chrome HT...



Ranking (Punteggio)

1. Selezione del file delle regole per lo specifico incarico
2. Selezione del set dei documenti da analizzare in formato testo **DOC/DOCX/ODT/PDF** (eventualmente salvati con titolo predefinito, nome candidato o numero, per favorire l'ordinamento)
 - Produzione automatica del file PUNTEGGIO.XLS (directory Windows DOWNLOAD)
 - Il sistema indica il numero di documenti presi in considerazione
 - Tempi di analisi rapidissimi (con dimensione media documenti: centinaia di KB): **pochi decimi di secondo a documento** anche con centinaia di regole (tempi più lunghi se interviene OCR)
 - Se un documento è illeggibile o non contiene testo, il punteggio è zero o molto basso (interviene OCR)
 - Funzione di default all'apertura dell'app

The screenshot displays the 'PBS Talent - Words' application interface. The main window shows a search path for 'CV LAN & WAN Sales' and a list of files. A secondary window shows the 'Download' folder with a file named 'Punteggio -30_04_2021-10_07_19' selected.

Nome	Modificato	Dimensioni	Tipo
i05	29/04/2021 17.38	63,3 KB	Chrome HT...
i04	29/04/2021 17.38	184 KB	Chrome HT...
i04	29/04/2021 09.26	258 KB	Chrome HT...
i03	29/04/2021 09.26	1,00 MB	Chrome HT...
i03	28/04/2021 10.31	332 KB	Chrome HT...
i02	28/04/2021 10.31	65,1 KB	Documento ...
i02	28/04/2021 10.27	218 KB	Chrome HT...
i01	28/04/2021 10.27	257 KB	Chrome HT...
i01	28/04/2021 10.23	265 KB	Chrome HT...
i01	28/04/2021 10.23	167 KB	Chrome HT...

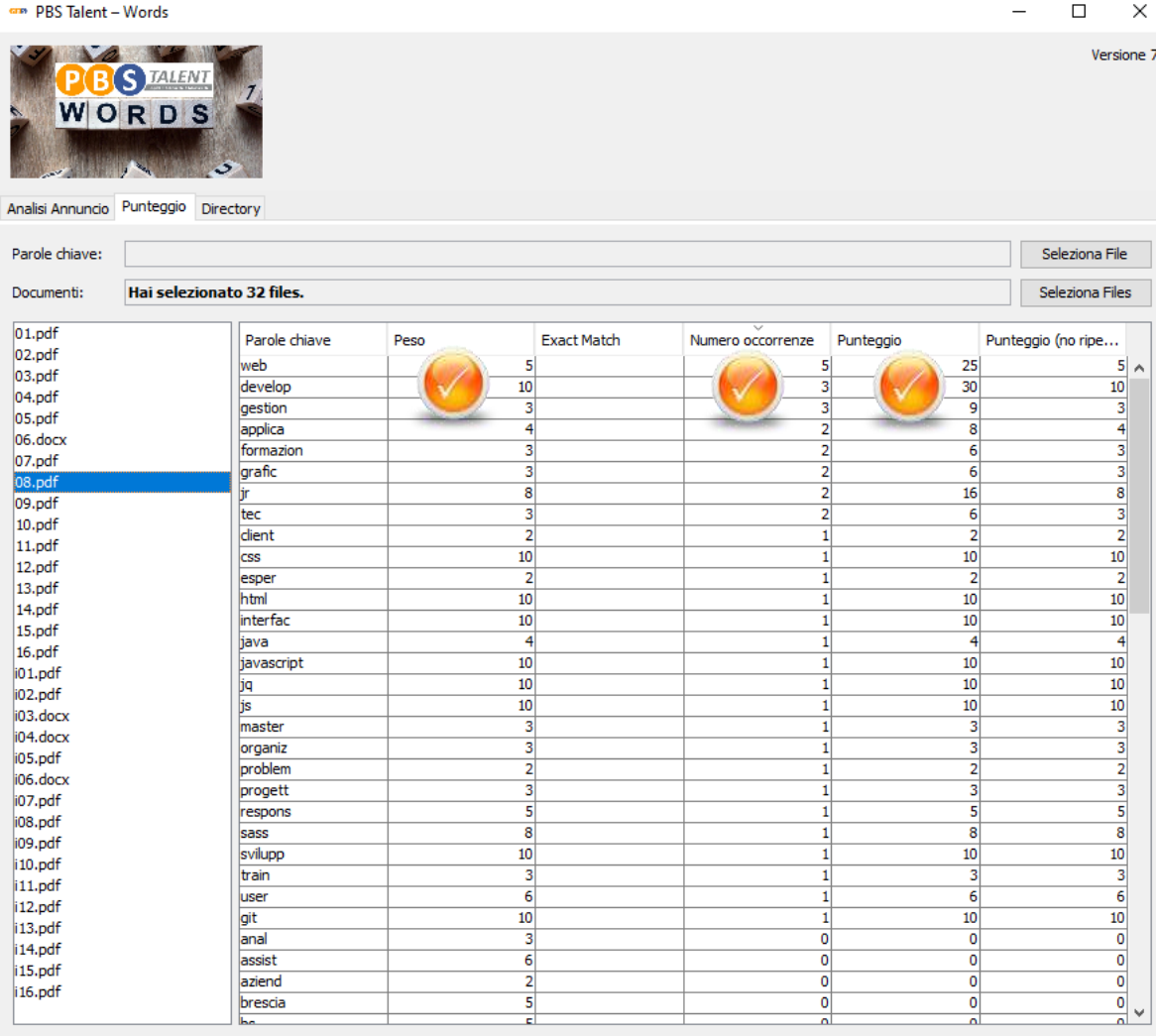
File XLS di ranking (Punteggio): 6 colonne

1. **Nome file:** in ordine alfabetico
2. **Punti:** somma dei punteggi delle regole riscontrate, incluse le ripetizioni (*se una regola vale 10 punti e viene trovata 5 volte, porta 50 punti; Se viene trovata 1 volta porta 10 punti*)
3. **Punti (no ripetizioni):** somma dei punteggi delle regole riscontrate dove ogni regola viene calcolata una sola volta; Utile per non venire influenzati da documenti verbosi (*se una regola vale 10 punti e viene trovata 5 volte oppure 1 volta, porta sempre 10 punti*)
4. **Numero parole:** del documento, per dare valore alla lunghezza del testo
5. **Numero caratteri:** del documento, per dare valore alla lunghezza del testo
6. **Inglese?:** «1» se il documento è in lingua inglese, «0» se no

Nome File	Punti	Punti (no ripetizioni)	Numero parole	Numero caratteri	Inglese?
01.pdf	209	96	464	3661	0
02.pdf	479	163	709	5262	0
03.pdf	509	139	1039	6579	1
04.pdf	1045	293	1070	7907	1
05.pdf	852	215	831	6350	0
06.pdf	1013	358	1008	7394	0
07.pdf	470	194	384	2929	1
08.pdf	1071	257	1086	8103	0
09.pdf	658	185	565	4610	1
10.pdf	181	84	543	4328	0

Il ranking di ogni documento può essere esplorato in dettaglio

1. Cliccare su un file
2. Scegliere la colonna su cui ordinare (crescente o decrescente)
3. Analizzare



The screenshot shows the 'PBS Talent - Words' application window. At the top, there is a logo with the text 'PBS TALENT WORDS'. Below the logo, there are tabs for 'Analisi Annuncio', 'Punteggio', and 'Directory'. A search bar labeled 'Parole chiave:' is present, along with a 'Seleziona File' button. Below the search bar, it says 'Documenti: Hai selezionato 32 files.' and another 'Seleziona Files' button. The main area contains a table with the following columns: 'Parole chiave', 'Peso', 'Exact Match', 'Numero occorrenze', 'Punteggio', and 'Punteggio (no ripe...'. The table lists various keywords and their associated metrics for 32 files. The file '08.pdf' is highlighted in blue. The 'Peso' column contains orange circular icons with a checkmark.

	Parole chiave	Peso	Exact Match	Numero occorrenze	Punteggio	Punteggio (no ripe...)
01.pdf	web	5		5	25	5
02.pdf	develop	10		3	30	10
03.pdf	gestion	3		3	9	3
04.pdf	applica	4		2	8	4
05.pdf	formazion	3		2	6	3
06.docx	grafic	3		2	6	3
07.pdf	jr	8		2	16	8
08.pdf	tec	3		2	6	3
09.pdf	client	2		1	2	2
10.pdf	css	10		1	10	10
11.pdf	esper	2		1	2	2
12.pdf	html	10		1	10	10
13.pdf	interfac	10		1	10	10
14.pdf	java	4		1	4	4
15.pdf	javascript	10		1	10	10
16.pdf	jq	10		1	10	10
i01.pdf	js	10		1	10	10
i02.pdf	master	3		1	3	3
i03.docx	organiz	3		1	3	3
i04.docx	problem	2		1	2	2
i05.pdf	progett	3		1	3	3
i06.docx	respons	5		1	5	5
i07.pdf	sass	8		1	8	8
i08.pdf	svilup	10		1	10	10
i09.pdf	train	3		1	3	3
i10.pdf	user	6		1	6	6
i11.pdf	git	10		1	10	10
i12.pdf	anal	3		0	0	0
i13.pdf	assist	6		0	0	0
i14.pdf	aziend	2		0	0	0
i15.pdf	brescia	5		0	0	0
i16.pdf	be	5		0	0	0

Ranking candidati, #1

SEMPLICE: Scegliere un criterio e ordinare su quello

Nome File	Punti	Punti (no ripetizioni)	Numero parole	Numero caratteri	Inglese?
08.pdf	1071	257	1086	8103	0
04.pdf	1045	293	1070	7907	1
06.pdf	1013	358	1008	7394	0
05.pdf	852	215	831	6350	0
09.pdf	658	185	565	4610	1
03.pdf	509	139	1039	6579	1
02.pdf	479	163	709	5262	0
07.pdf	470	194	384	2929	1
01.pdf	209	96	464	3661	0
10.pdf	181	84	543	4328	0

Ranking candidati, #2

SEMPLICE e NORMALIZZATO: Scegliere un criterio e ordinare su quello, normalizzando a 100

Nome File	Punti	Punti (no ripetizioni)	Numero parole	Numero caratteri	Inglese?	Punti %
08.pdf	1071	257	1086	8103	0	100%
04.pdf	1045	293	1070	7907	1	98%
06.pdf	1013	358	1008	7394	0	95%
05.pdf	852	215	831	6350	0	80%
09.pdf	658	185	565	4610	1	61%
03.pdf	509	139	1039	6579	1	48%
02.pdf	479	163	709	5262	0	45%
07.pdf	470	194	384	2929	1	44%
01.pdf	209	96	464	3661	0	20%
10.pdf	181	84	543	4328	0	17%



Ranking, #3

SOFISTICATO A PIACERE:

- Dare peso arbitrario (*criterio di management*), incarico per incarico, a tutte o alcune colonne
- Calcolare il punteggio totale (sommatoria «*valore * peso criterio*»)
- (opzionale) Normalizzare a 100

Nome File	Punti	Punti (no ripetizioni)	Numero parole	Numero caratteri	Inglese?	Rank	Punti %
08.pdf	1071	257	1086	8103	0	60883	100%
04.pdf	1045	293	1070	7907	1	59825	98%
06.pdf	1013	358	1008	7394	0	57020	94%
05.pdf	852	215	831	6350	0	47807	79%
03.pdf	509	139	1039	6579	1	46376	76%
02.pdf	479	163	709	5262	0	37367	61%
09.pdf	658	185	565	4610	1	35071	58%
10.pdf	181	84	543	4328	0	27923	46%
01.pdf	209	96	464	3661	0	24411	40%
07.pdf	470	194	384	2929	1	23591	39%
Peso	10	8	7	5	6		

Successiva ottimizzazione database globale delle regole: inserimento nuove regole, eliminazione ridondanze

0	Exact_Match	Additional	Additional #2	Punteggio base		Duplicate?	
68				10	1	0%	0
365				5	2	0%	0
365				5	3	0%	0
730				5	4	0%	0
740				5	5	0%	0
770				5	6	0%	0
802				5	7	0%	0
900				5	8	0%	0
2d				3	9	0%	0
3cx				5	10	0%	0
3d				8	11	1%	0
3ntr				10	12	1%	0
a10				8	13	1%	0
a4				5	14	1%	0
abb	yes			5	15	1%	1
abb				3	16	1%	2

- Copia nuove regole nel database globale
- Eliminazione dei duplicati
- Verifica Exact_Match di default

Licensing

1. **Licenza permanente:** uso software per azienda e sue consociate
 - **Avviamento:** fornitura database globale delle regole, «training on the job» per 2 annunci, fornitura di modelli precompilati di regole (Developer, Sales, IT Engineer ...), supporto tecnico (30 giorni)
2. **Abbonamento annuo** a nuove versioni (mediamente 1-2/anno) + supporto tecnico (annuo)
3. **Fruizione come Servizio:** dato annuncio/documenti/criteri di ranking, incarico per incarico, viene creato un database di regole specifiche e fornito il ranking secondo i criteri concordati
4. **Altri sviluppi software:** quotazione ad hoc per nuove funzioni da inserire a sistema o «fork»
5. **Fornitura sorgenti:** sviluppi ad hoc da parte del cliente
6. **Contatto software house:** sviluppi ad hoc



Algoritmi semantici a supporto del Digital Recruiting: Proviamoli insieme! (www.primobonacina.com/words)

