

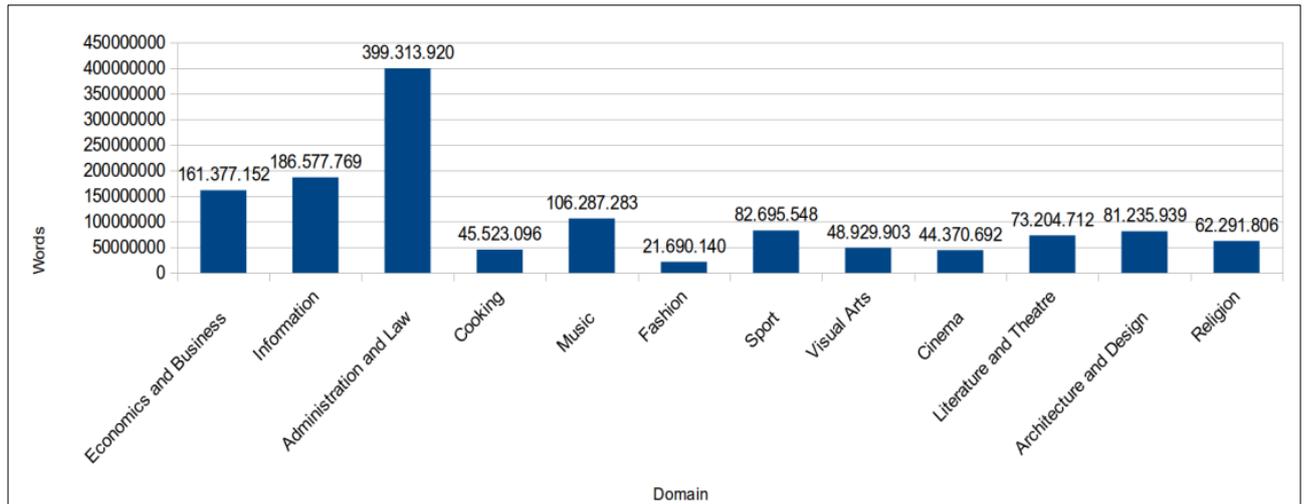
FUNZIONI DI RICERCA NEL WEB CORPUS RIDIRE - MANUALE UTENTE

Massimo Moneglia (Università di Firenze)

1 Il Corpus Ridire

Il corpus RIDIRE contenente circa 1.5 miliardi di tokens estratti dal web italiano tra il 2009 e il 2013. Il corpus contiene risorse divise per dominio, rispettivamente domini definiti dal loro carattere semantico o dalla loro funzione ed è composto come riportato nella tabella e negli istogrammi di seguito riportati

DOMINI	# WEBSITES	# PAGINE	# TOKENS	# PAROLE
Funzionali (totale)	189	976.460	854.388.230	747.268.841
Informazione	27	550.169	216.431.868	186.577.769
Economia & Business	123	226.535	179.710.476	161.377.152
Amministrazione & Legislazione	39	199.756	458.245.886	399.313.920
Semantici (totale)				
Sport	49	138.235	98.172.470	82.695.548
Architettura & Design	142	136.725	93.822.675	81.235.939
Cucina	20	123.376	52.784.045	45.523.096
Cinema	25	122.850	51.466.145	44.370.692
Musica	195	113.015	12.906.213	106.287.283
Moda	103	74.584	24.645.980	21.690.140
Arti Visive	118	70.601	56.517.442	48.929.903
Religione	51	66.053	72.454.492	62,291.806
Ltteratura & Teatro	113	61.935	85.474.102	73.204.712
Totale	2.010	3.767.668	1.514.631.794	1.313.497.960



2 Interfaccia di query

L'utente accede liberamente alle funzioni di ricerca della area di query senza registrazione e potrà selezionare le funzioni disponibili, ovvero *liste di frequenza*, *concordanze*, *patterns* e *sketches*.

Il log-in è riservato alle funzioni di gestione, monitoraggio e up-date del corpus.

RIDIRE ti dà il benvenuto

- [Entra](#)
- Querying Area**
 - [Liste di frequenza](#)
 - [Concordanze](#)
 - [Pattern](#)
- Sketch Area**
 - [Lemma singolo](#)
 - [Confronto due lemmi](#)
 - [Confronto due domini](#)

[Documenti](#) [Contatti](#) [Progetto](#) [Area riservata](#)

3. Liste di frequenza

Le liste di frequenza sono fornite per lemma per tutto il corpus (tutti i domini), o di ciascun sottocorpus separatamente (all'interno delle serie funzionali o semantiche), sono pre-calcolate e possono essere scaricate dall'utente sul proprio computer in un file open office (ods), apribile anche con excell.

Liste di frequenza

Domini funzionali

Dominio	Dimensione (byte)	Scarica
Amministrazione_e_Legislazione	234167	↓
Informazione	407152	↓

Domini semantici

Dominio	Dimensione (byte)	Scarica
Architettura_e_Design	269742	↓
Arti_figurative	360341	↓
Cinema	360341	↓
Cucina	138175	↓
Letteratura_e_Teatro	342015	↓
Moda	261314	↓
Musica	227253	↓
Religione	270381	↓
Sport	259888	↓

Tutti i domini

Dominio	Dimensione (byte)	Scarica
TOTALE	491485	↓

Tutte le liste di frequenza

Dominio	Dimensione (byte)	Scarica

Ciascuna lista può essere scaricata separatamente o tutte le liste possono essere scaricate in un solo pacchetto (tutte le liste di frequenza).

Il file scaricato contiene per ciascun lemma il numero di occorrenze, e la *parte del discorso* (POS) assegnata con *tree tagger* (Tagset-Baroni semplificato). Le liste dei tagset (esteso e versione semplificata) sono qui riportate in figura di seguito.

Le liste di frequenza sono quindi ordinabili per POS usando le normali funzioni di ordinamento del foglio elettronico.

I lemmi non riconosciuti sono assegnati da treetagger a lemmi “unknown” a cui è comunque assegnata una PoS (ci sono quindi tanti lemmi unknown quante sono le pos). Tutte le forme della stessa pos non riconosciute sono assegnate allo stesso lemma “unknown”.

La figura seguente mostra come si presentano le liste di frequenza, rispettivamente ordinate per frequenza assoluta (di default) o per POS.

	A	B	C	D
1	24758425	<unknown>	NPR	
2	16695105	di	PRE	
3	11991685	e	CON	
4	11518612	<unknown>	NOUN	
5	11456363	il	ART	
6	11084314	<unknown>	ADJ	
7	9485793	la	ART	
8	8626231	del	ARTPRE	
9	6966790	<unknown>	VER	
10	6734525	<unknown>	NOCAT	
11	6296464	in	PRE	
12	6142698	che	CHE	
13	5841977	a	PRE	
14	5081995	per	PRE	
15	4897184	della	ARTPRE	
16	4638075	un	ART	
17	4558190	essere	VER	
18	4235137	avere	VER	
19	4117688	al	ARTPRE	
20	3370506	con	PRE	
21	3200835	il	PROdef	
22	3103111	non	NEG	
23	3053283	una	ART	
24	3030892	l'	ART	
25	2661959	si	CLI	
26	2658005	nel	ARTPRE	
27	2647383	da	PRE	
28	2128404	alla	ARTPRE	

A1		fx 94627	
A	B	C	
33	207171	presente	ADJ
34	205673	tale	ADJ
35	180931	secondo	ADJ
36	171376	importante	ADJ
37	155432	possibile	ADJ
38	154873	prossimo	ADJ
39	151523	unico	ADJ
40	150938	piccolo	ADJ
41	146911	internazionale	ADJ
42	142298	umano	ADJ
43	134687	nazionale	ADJ
44	131995	scorso	ADJ
45	130510	relativo	ADJ
46	129267	sociale	ADJ
47	128315	bello	ADJ
48	126195	europeo	ADJ
49	126109	forte	ADJ
50	124299	alto	ADJ
51	121861	lungo	ADJ
52	120647	necessario	ADJ
53	116922	previsto	ADJ
54	114803	politico	ADJ
55	110840	giovane	ADJ
56	108617	certo	ADJ
57	108119	economico	ADJ
58	108103	buono	ADJ
59	107105	solo	ADJ
60	105013	generale	ADJ

ADJ	adjective
ADJPRO	pronominal adject
ADV	adverb (excluding
ADV:mente	adverb ending in -
ART	article
CHE	che
CLI	clitic
CON	conjunction
NEG	negation
NOCAT	non-linguistic el
NOUN	noun
NPR	proper noun
NUM	number
PREP	preposition
PRON	pronoun
PUN	non-sentence-fin
SENT	sentence-final pu
VERB	verb
WH	wh word

Tagset semplificato

ADJ	adjective
ADV	adverb (excluding -mente forms)
ADV:mente	adverb ending in -mente
ART	article
ARTPRE	preposition + article
AUX:fin	finite form of auxiliary
AUX:fin:cli	finite form of auxiliary with clitic
AUX:geru	gerundive form of auxiliary
AUX:geru:cli	gerundive form of auxiliary with clitic
AUX:infi	infinitival form of auxiliary
AUX:infi:cli	infinitival form of auxiliary with clitic
AUX:ppast	past participle of auxiliary
AUX:ppre	present participle of auxiliary
CHE	che
CLI	clitic
CON	conjunction
DET:demo	demonstrative determiner
DET:indef	indefinite determiner
DET:num	numeral determiner
DET:poss	possessive determiner
DET:wh	wh determiner
NEG	negation
NOCAT	non-linguistic element
NOUN	noun
NPR	proper noun
NUM	number
PRE	preposition
PRO:demo	demonstrative pronoun
PRO:indef	indefinite pronoun
PRO:num	numeral pronoun
PRO:pers	personal pronoun
PRO:poss	possessive pronoun
PUN	non-sentence-final punctuation mark
SENT	sentence-final punctuation mark
VER2:fin	finite form of modal/causal verb
VER2:fin:cli	finite form of modal/causal verb with c
VER2:geru	gerundive form of modal/causal verb
VER2:geru:cli	gerundive form of modal/causal verb with c
VER2:infi	infinitival form of modal/causal verb
VER2:infi:cli	infinitival form of modal/causal verb with c
VER2:ppast	past participle of modal/causal verb
VER2:ppre	present participle of modal/causal verb
VER:fin	finite form of verb
VER:fin:cli	finite form of verb with clitic
VER:geru	gerundive form of verb
VER:geru:cli	gerundive form of verb with clitic
VER:infi	infinitival form of verb
VER:infi:cli	infinitival form of verb with clitic
VER:ppast	past participle of verb
VER:ppast:cli	past participle of verb with clitic
VER:ppre	present participle of verb
WH	wh word

Tagset esteso

4. Concordanze e collocazioni

4.1 Concordanza semplice

Le concordanze possono essere fornite alternativamente per forma, per lemma e per phrase (i phrase, anche le polirematiche, non sono lemmatizzati per es. “conto corrente” è un phrase, la loro ricerca è quindi più lenta)

Le concordanze per forma e per lemma possono essere ristrette per pos. Le due serie di pos, estesa e ristretta, sono dettagliate a lato.

NELLE QUERY IN CUI SI SPECIFICA UN POS LE SCELTE RELATIVE AL TAGSET SEMPLIFICATO SONO IL DEFAULT E COMPAGNONO IN UN MENU' A TENDINA. Se l'utente vuole usare il tagset esteso, il tag deve essere digitato. I tagset sono sempre riportati a latere della pagina delle query, come nella figura seguente.

Le concordanze possono essere estratte dall'intero corpus (tutti i domini) e questa è la scelta di default. Altrimenti possono essere estratte su un dominio selezionato dopo aver scelto l'opzione “No” e messo la spunta sul dominio che interessa.¹ L'esempio seguente mostra la ricerca della forma verbale “uscita” nel solo sottocorpus “informazione”

Concordanze semplici

Forma: Lemma: PoS:

PoS estesi: Phrase:

Tutti i domini: Sì No

Scegli dominio semantico

- Architettura e design
- Arti figurative
- Cinema
- Cucina
- Letteratura e Teatro
- Moda
- Musica
- Religione
- Sport

Scegli dominio funzionale

- Amministrazione e Legislazione
- Economia e Affari
- Informazione

Tagset	
ADJ	adjective
ADJPRO	pronominal
ADV	adverb (e)
ADV:mente	adverb endi
ART	article
CHE	che
CLI	clitic
CON	conjunctio
NEG	negation
NOCAT	non-lingui
NOUN	noun
NPR	proper nou
NUM	number
PREP	prepositio
PRON	pronoun
PUN	non-senter
SENT	sentence-f
VERB	verb
WH	wh word

Tagset esteso

I risultati delle concordanze sono completi, ovvero tutte le occorrenze nel corpus selezionato sono riportate, e sono visibili scorrendo le pagine.

¹ Si possono fare ricerche di concordanze su sottoinsiemi di domini (opzione teorica non significativa per ridire), ma in questo caso l'utente non potrà estrarre le collocate in quanto i risultati non sono indicizzati

Ordinamento

Nessun ordine
 Testo cercato
 Contesto sinistro
 Contesto destro

Contesto per ordinamento:

Raggruppamento a destra

Sì/No N. forme:
 Raggruppa per lemma forma

Cerca Reset Estrai collocate

[Genera tabella concordanze](#)

[« Prima pagina](#)
[« Precedente](#)
[Prossima »](#)
[Ultima pagina »»](#)

Risultati: 1 - 20 di 1063

Sx		Dx	Dominio	URL
l'ultimo Spagna-Italia 0-0 (4-2) L' Italia è	uscita	per l' errore di Pagliuca contro il Brasile di quattordici	Informazione	
il 6,09 % . Unicredit nel finale è entrata e	uscita	dall' asta di volatilità con un ribasso teorico del 4,34	Informazione	
lo stop a Castro Pretorio - commenta infuriata una signora	uscita	dalla metropolitana - lo arrivo dall' Eur e sono già	Informazione	
ride . Ormai è più sciolta . Come se fosse	uscita	dai postumi di un' anestesia . « Figurati . Dobbiamo	Informazione	
« Veronica , mi fa piacere che da lunedì è	uscita	la vera donna che c'è in te , una guerriera	Informazione	
a nomi , cose , città , animali . Era	uscita	la " i " . Alla voce " cose "	Informazione	
scrittore e guru dei blogger statunitensi , parla della canzone	uscita	nel 1972 in toni entusiastici , la cita come esempio	Informazione	

I risultati permettono l'accesso tramite link alla pagina web sorgente (url), così come si presenta alla data della ricerca, quindi l'utente deve tener conto che il contenuto in rete può essere stato ma solo limitatamente a una citazione (intorno di massimo 100 parole precedenti e seguenti la parola cercata).

Ordinamento

Nessun ordine
 Testo cercato
 Contesto sinistro
 Contesto destro

Contesto per ordinamento:

Raggruppamento a destra

Sì/No N. forme:
 Raggruppa per lemma forma

Cerca Reset

[Genera tabella con](#)

« Prima pagina

Risultati: 1 - 20 di

Contesto

con cui i turchi arrivano a Vienna Olanda-Romania 2-0 " Hannibal the Cannibal era rumeno " Italia-Francia 2-0 Nel salotto del cinico tutti festeggiano un' Italia con l' oroscopo a favore e Donadoni ha imparato Russia-Svezia 2-0 Germania-Portogallo 3-2 La sindrome di Stoccolma dei tedeschi e i baffi di Scolari e Cristiano Ronaldo e la Questione Antropologica Turchia-Croazia 1-1 La Turchia ha una rabbia calcistica che il Vecchio Continente si sogna Germania-Turchia 3-2 Il black out non nasconde il contrappasso di Terim e Tutta la Turchia minuto per minuto , tranne l' ultimo Spagna-Italia 0-0 (4-2) L' Italia è uscita per l' errore di Pagliuca contro il Brasile di quattordici anni fa ... © - FOGLIO QUOTIDIANO di Maurizio Crippa PAOLO VI ANGELUS DOMINI Solennità dell' Epifania di Nostro Signore Lunedì , 6 gennaio 1969 È l' Epifania ! Oggi il Nostro pensiero , e per ciò la Nostra preghiera , va alla intera umanità chiamata alla luce del Vangelo . L' Epifania è la festa per i Popoli , è la festa delle Missioni , è la festa del mondo al quale è offerta una vocazione alla verità su Dio , su Cristo , su la Chiesa , su la

Dominio	URL
Informazione	

Lido tre buchi : la vittima era appena uscita dall' acqua quando si è accasciata sulla spiaggia .

E' anche possibile generare la tabella excell delle concordanze, ma questa utilità, che può essere utilizzata per fini di ricerca, è molto lenta (nell'ordine delle decine di minuti) e non può assolutamente essere usata

per ricerche on the fly, che sono lo scopo primario di RIDIRE. L'utente può monitorare la percentuale di tabella generata cliccando su *refresh*.

4.2 Ordinamento dei risultati delle concordanze

L'infrastruttura prevede tre livelli di ordinamento dei risultati:

- 1- Per contesto destro o sinistro
- 2- Per raggruppamento in frequenza
- 3- Per collocate

Per ottenere qualsiasi tipo di ordinamento deve sempre essere prima estratta la concordanza. Per esempio non si possono chiedere le collocate di un lemma direttamente, ma si richiederà questo tipo di ordinamento dopo che è stata generata la concordanza.

Le ricerche che danno luogo ai diversi tipi di ordinamento di una concordanza possono essere fatte in serie, ovvero senza cercare di nuovo la concordanza, semplicemente scegliendo i nuovi parametri di ordinamento e cliccando su "cerca".

4.2.1 Ordinamento per contesto

L'ordinamento per contesto sinistro o destro prevede la definizione della distanza del testo cercato dell'elemento destro o sinistro (l'ordinamento per il primo elemento a destra o a sinistra è di default).

L'utente deve considerare che nel numero sono contati anche i tokens che non sono parole ; per es. (, " etc.)

Di seguito gli ordinamenti della ricerca precedente per il primo elemento a sinistra:

Ordinamento

Nessun ordine
 Testo cercato
 Contesto sinistro
 Contesto destro

Contesto per ordinamento:

Raggruppamento a destra

Si/No N. forme:
 Raggruppa per lemma forma

70 % [Refresh](#)

Risultati: 1 - 20 di 1063

Sx		Dx	Dominio	URL
seconda edizione dei " Documenti , fonti , inventari "	uscita	appena nel 2010. Ma se i documenti nuovi sono "	Informazione	
per indulto di tre anni . La decisione e '	uscita	oggi pomeriggio dopo quasi tre ore di camera di consiglio	Informazione	
per indulto di tre anni . La decisione e '	uscita	oggi pomeriggio dopo quasi tre ore di camera di consiglio	Informazione	
poco dopo le due di pomeriggio e ne e '	uscita	poco dopo le 18. Il verbale di interrogatorio e '	Informazione	
' sempre proclamata innocente , dopo la sentenza e '	uscita	piangendo dall' aula e si e ' stretta fra le	Informazione	
dalla sua gemella siamese nata senza testa , e '	uscita	oggi dall' ospedale ed e ' tornata a casa con	Informazione	
" . Nel frattempo , pero ' , e '	uscita	la notizia della nuova accusa e il primo cittadino all'	Informazione	
' senso della misura " . L' Italia e '	uscita	bene dal G8 L' Italia " nell' insieme e '	Informazione	
' forti , oltre a Busi , ieri e '	uscita	anche Loredana Lecciso . Inserire corpo " Il direttore di	Informazione	

Di seguito un esempio di ordinamento per il secondo token a destra

Ordinamento

Nessun ordine
 Testo cercato
 Contesto sinistro
 Contesto destro

Contesto per ordinamento:

Raggruppamento a destra

Sì/No N. forme:
 Raggruppa per lemma forma

[Cerca](#) [Reset](#) [Estrai collocate](#)

[Genera tabella concordanze](#)

[« Prima pagina](#)
[« Precedente](#)
[Prossima »](#)
[Ultima pagina »](#)

Risultati: **1 - 20** di **1063**

Sx		Dx	Dominio	URL
la Aniston , che ha raccontato ai tabloid di essere	uscita	" devastata " dal divorzio , avrà bisogno di tempo	Informazione	
la Aniston , che ha raccontato ai tabloid di essere	uscita	" devastata " dal divorzio , avrà bisogno di tempo	Informazione	
del tessuto sociale " . Secondo Bayrou , Royal è	uscita	" piuttosto bene " dal dibattito televisivo . Al primo	Informazione	
del tessuto sociale " . Secondo Bayrou , Royal è	uscita	" piuttosto bene " dal dibattito televisivo . Al primo	Informazione	
buona : stava dormendo , non stava bene , era	uscita	' ' . "lo sono riuscita a vederla l' ultima	Informazione	
diventato un concertista » : così , nell' autobiografia recentemente	uscita	, Lang Lang ricorda il momento clou della sua infanzia	Informazione	
da morire , non potevo neanche alzarmi , ma sono	uscita	, avevamo una riunione al ministero degli Affari esteri ,	Informazione	
mai preoccupato delle traduzioni , e per un secolo è	uscita	, con pochissime varianti , sempre la stessa . Donzelli	Informazione	

4.2.2. Ordinamento per raggruppamento

L'ordinamento per raggruppamento serve a dare una prima misura puramente **percentuale**, ma a colpo d'occhio significativa, delle associazioni di una parola nella concordanza estratta.

La ricerca si ottiene spuntando il box e cliccando su *cerca*

L'associazione misurata è solo con i tokens a destra (più significativo in italiano) e si può specificare la distanza in tokens sui quali si vuole raggruppare. Il default è distanza 1.

Il raggruppamento può essere misurato sulle forme o, più significativamente, per lemma.

Il seguente è il calcolo dei raggruppamenti in ordine di frequenza della forma verbale "uscita" rispetto al primo lemma (sulla destra), con risultato interessante dal punto di vista sintattico: "uscita da", ma anche "uscita in", "uscita la" ecc. con percentuali significative, indicate sulla destra:

Raggruppamento a destra

Si/No N. forme: 1 Raggruppa per lemma forma

Cerca Reset Estrai collocate

[Genera tabella concordanze](#)

« Prima pagina < Precedente Prossima > Ultima pagina »

Risultati: 1 - 20 di 117

Num.	Sx	Testo	Dx	Dominio	URL
+ 144 13,55%	, 16-04-2010 Da ieri mattina Sandra Mondaini non è più	uscita dalla	sua abitazione , la residenza Acquario a Segrate Milano 2.	Informazione	
+ 129 12,14%	davanti all' ingresso di corso Allamano . L' ultima auto	uscita dai	cancelli di Grugliasco è del 2006. Anche per lui sono	Informazione	
+ 111 10,44%	(13.43) , mentre la statunitense Nia Ali è	uscita di	gara in anticipo infortunandosi al settimo ostacolo . Piano traffico	Informazione	
+ 65 6,11%	22 maggio al 17 ottobre . Il 6 giugno è	uscita la	" Conversione di Saulo " della collezione Odescalchi . Viceversa	Informazione	
+ 60 5,64%	15-10-2003 La navetta spaziale cinese " Shenzhou 5 " è	uscita dall'	orbita ellittica ed è entrata in quella circolare intorno alla	Informazione	
+ 55 5,17%	a Buona domenica ? - ha dunque scelto ed è	uscita da	quella bolgia senza capo né coda , senza notte né	Informazione	
+ 33 3,1%	29 settembre . « Adriana , quel pomeriggio , era	uscita con	il fidanzato e la sua amica : una ragazza brasiliana	Informazione	
+ 31 2,92%	alcuni rinvii , sempre alla fine del 2000 , è	uscita una	norma dove si stabilisce che i bolli sui rimorchi vengono	Informazione	
+ 24 2,26%	ai genitori . Verso le ore 19:00 , dopo essere	uscita in	anticipo dalla lezione , telefonò a casa per riferire la	Informazione	
+ 22 2,07%	quanto leader che crea il sisma . La Spagna è	uscita ai	Mondiali quando lui ha deciso di renderle omaggio , Ségolène	Informazione	
+ 20 1,88%	cominciato ieri le prime scaramucce . Con sospetto tempismo è	uscita sui	media inglesi la notizia che Adrian Newey , geniale progettista	Informazione	

Il seguente è il raggruppamento della forma verbale “uscita” sui due tokens seguenti, con risultato più interessante dal punto di vista dell’associazione lessicale

Raggruppamento a destra

Si/No N. forme: 2 Raggruppa per lemma forma

Cerca Reset Estrai collocate

[Genera tabella concordanze](#)

« Prima pagina < Precedente Prossima > Ultima pagina »

Risultati: 1 - 20 di 516

Num.	Sx	Testo	Dx
+ 42 3,95%	26 agosto scorso per andare al mare . Sarah è	uscita di casa	alle 14,30 per raggiungerla a casa , a poche centinaia
+ 31 2,92%	a bordo una famiglia di Asti ha sbandato ed è	uscita di strada	ribaltandosi nella scarpata . Nell' impatto hanno perso la vita
+ 19 1,79%	Ajmi : " For peace " . Sempre ieri è	uscita la notizia	che Mohammed al Qatani , uno degli aspiranti direttori d
+ 14 1,32%	condanna per guida in stato di ebbrezza . Era poi	uscita dal carcere	femminile di Lynwood dopo appena 13 giorni di galera graz
+ 13 1,22%	Una colonna di circa mille mezzi della Guardia repubblicana è	uscita da Baghdad	diretta a sud , verso Najaf , il teatro nelle
+ 13 1,22%	una forza incredibile . Due giorni dopo lo stupro è	uscita con gli	amici per le vie della città , senza timore delle
+ 13 1,22%	al lavoro per presentare il progetto entro 10 giorni L'	uscita allo scoperto	di Colaninno ieri sul piano Fiat forse non è stata
+ 11 1,03%	è garantita Marilena Piretti , 27enne di Trieste , è	uscita da casa	il 14 gennaio . Da allora non si hanno più
+ 10 0,94%	. A quanto si apprende , la piccola era appena	uscita da un	istituto scolastico quando e ' stata avvicinata e caricata in
+ 10 0,94%	di voti parlamentari) spezza simbolicamente l' autosufficienza della maggioranza	uscita dalle urne	del 2008 " , fa notare , fra gli altri
+ 9 0,85%	azzurre in gara nella sessione mattutina : Raffaella Lamera è	uscita di scena	nell' alto con 1,85 alla terza e tre errori ad
+ 9 0,85%	la notte " . La Ferrari del brasiliano ieri era	uscita di pista	al termine della seconda manche delle qualifiche del Gran

4.2.3 Ordinamento per collocazioni

L'estrazione delle collocate si richiede, come gli altri ordinamenti, di aver estratto una concordanza.

Il menù è più complesso.

Collocate

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Si può scegliere la misura statistica. Il default è la logDice, che corregge vari problemi della *mutual information*.

Le collocate possono essere ristrette in relazione alla frequenza minima, perché la misura statistica possa essere considerata a fini linguistici.

La frequenza minima riguarda sia dell'associazione tra espressioni nella concordanza ("frequenza minima correlazione") sia la frequenza assoluta della parola che colloca con la parola cercata nelle concordanze (frequenza minima parola collocata).

Il default (5) evita i problemi maggiori degli algoritmi, ma può essere alzato o abbassato in relazione al numero di concordanze ottenute, in particolare, se le concordanze sono molte migliaia, può convenire aumentare la soglia.

La collocata può riguardare sia una forma sia un lemma.

Crucialmente l'utente deve decidere il numero di parole a sinistra e a destra che debbono essere considerate per l'associazione. Tale numero dipende dal tipo di associazione che si vuol cercare.

La seguente sono le collocazioni generiche della forma "uscita" nel sottocorpus informazione. Come si vede sono individuate collocazioni sia a destra che a sinistra di "Uscita" nel suo intorno

Collocate

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Estrai collocate

« Prima pagina « Precedente **Prossima** » Ultima pagina »

Risultati: 1 - 20 di 1187

Parola	Punteggio	Freq. nel corpus	Freq. collocata	Freq. collocata attesa
banarotta	0,0112	9	6	1.0E-4
palla	0,008	696	7	0.0092
dalla	0,0078	36094	144	0.4786
Lubiana	0,0052	1264	6	0.0168
riformista	0,0051	1705	7	0.0226
ambulanza	0,0046	1991	7	0.0264
repubblicana	0,0039	2017	6	0.0267
indenne	0,0034	1851	5	0.0245
finalmente	0,0032	5270	10	0.0699
urne	0,0032	6411	12	0.085
ragazzina	0,0031	4177	8	0.0554
clinica	0,0027	5710	9	0.0757

(escludendo il campo punteggio e il campo frequenza collocata attesa, che non sono significativi per l'utente) il campo "frequenza nel corpus" (della parola associata) e "frequenza collocata" (ovvero dell'associazione), danno la misura della rilevanza dell'associazione. Per es. La collocata è importante se mi interrogo sull'associazione di "banarotta" che si associa a uscita 6 volte sulle 9 volte in cui banarotta occorre.

La seguente ricerca esclude gli elementi a destra dal calcolo ed è più orientata alla ricerca degli elementi che in italiano stanno prima del verbo (soggetti o ausiliari o avverbi).

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Estrai collocate

« Prima pagina < Precedente **Prossima** > Ultima pagina »

Risultati: 1 - 20 di 644

Parola	Punteggio	Freq. nel corpus	Freq. collocata	Freq. collocata attesa
palla	0,008	696	7	0.0046
Lubiana	0,0052	1264	6	0.0084
riformista	0,0051	1705	7	0.0113
ambulanza	0,0039	1991	6	0.0132
repubblicana	0,0039	2017	6	0.0134
ragazzina	0,0031	4177	8	0.0277
finalmente	0,0028	5270	9	0.0349
mattina	0,0012	7626	5	0.0506
appena	0,0012	76613	45	0.5079
era	0,0007	452472	170	2.9997
Guardia	0,0007	13180	5	0.0874
maggioranza	0,0006	51573	16	0.3419
sarebbe	0,0006	142640	42	0.9456

Un rilievo particolare nello sfruttamento delle collocate è la possibilità di restringere l'associazione ai soli elementi grammaticali sui cui l'utente si interroga.

La seguente ricerca relativa al lemma "uscire" evidenzia l'ambito associativo proprio delle proposizioni nel corpus cinema e per questo è ristretta ai soli elementi sulla destra che hanno POS Preposizione o preposizione articolata

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Estrai collocate

« Prima pagina < Precedente **Prossima** > Ultima pagina »

Risultati: **1 - 20 di 2669**

Parola	Punteggio	Freq. nel corpus	Freq. collocata	Freq. collocata attesa
dalla	0,0021	1526115	1602	196.7563
pro	0,0018	12930	32	1.667
dal	0,0017	2180045	1842	281.0651
nella	0,0015	1962616	1490	253.0327
nel	0,0012	3616567	2107	466.2705
dopo	0,0007	578009	200	74.5205
col	0,0007	130667	50	16.8464
in	0,0007	9002996	2956	1160.7226
da	0,0006	3581107	1050	461.6987
con	0,0005	4546696	1036	586.1885
dietro	0,0005	65167	21	8.4017
senza	0,0004	473170	105	61.004
fuori	0,0003	140584	24	18.125
entro	0,0003	186363	29	24.0271

La seguente ricerca evidenzia l'ambito associativo di uscire a destra (soggetti) sempre nel corpus Cinema

Collocate

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Estrai collocate

« Prima pagina < Precedente **Prossima** > Ultima pagina »

Risultati: **1 - 20 di 3111**

Parola	Punteggio	Freq. nel corpus	Freq. collocata	Freq. collocata attesa
pellicola	0,011	32853	303	4.2356
film	0,0099	405842	2111	52.3237
trailer	0,0028	6572	40	0.8473
frattempo	0,0024	17980	49	2.3181
poliziotto	0,0023	14499	42	1.8693
documentario	0,002	27745	49	3.5771
lungometraggio	0,0017	8035	25	1.0359
saga	0,0016	10242	26	1.3205
cinema	0,0012	136748	99	17.6304
ragazza	0,0011	56596	42	7.2967
personaggio	0,0011	125863	78	16.227
cauzione	0,001	7096	15	0.9149

Tale funzione, che è sfruttata in modo sistematico per il calcolo degli *sketches* (vedi oltre) , può essere sfruttata intensivamente per lo studio differenziale delle associazioni nei vari sottocorpora (sketch difference).

Per esempio le seguenti sono le associazioni con nomi a destra del lemma “uscire” nel corpus “sport”

Collocate

Distanza sinistra: Distanza destra:

Freq. minima correlazione: Freq. minima parola collocata:

Ordina: logDice MI Score Log Likelihood T Score

Basate su: Filtra su PoS:

Estrai collocate

« Prima pagina < Precedente **Prossima** > Ultima pagina »

Risultati: 1 - 20 di 2926

Parola	Punteggio	Freq. nel corpus	Freq. collocata	Freq. collocata attesa
palla	0,0036	21961	62	1.6116
tricolore	0,0028	9711	31	0.7126
pallone	0,0022	14690	30	1.078
giocatore	0,002	66753	79	4.8986
campione	0,0015	67841	60	4.9785
campionessa	0,0014	6584	13	0.4832
manche	0,0012	2418	9	0.1774
rigore	0,0012	26449	24	1.9409
tiro	0,0012	15240	16	1.1184
squadra	0,0012	155115	97	11.383
curva	0,0011	11469	13	0.8416
azzurra	0,0011	11598	13	0.8511

Dalle collocate si può sempre individuare i contesti di concordanza a cui ciascuna collocata si riferisce, cliccando sul termine collocato in azzurro. La seguente è la concordanza della collocata “squadra” associata a uscire.

[Genera tabella concordanze](#)

« Prima pagina « Precedente **Prossima** » Ultima pagina »

Risultati: 1 - 20 di 121

Sx		Dx	Dominio	URL
fa - racconta - fortunatamente io e due compagni di	squadra siamo usciti	illesi da uno scontro in auto . Quando vedi la	Sport	
il grande pubblico di fede granata- che possono aiutare la	squadra a uscire	da un momento di difficoltà e di sfiducia " .	Sport	
punto del campionato . Ma è altrettanto vero che la	squadra di Pellegrini era uscita	al primo turno della Coppa del Re . Ed è	Sport	
passare il turno . Poche saranno le possibilità per la	squadre che uscirà	perdente : troppo pochi i gol segnati per poter sperare	Sport	
nella terza partita del girone contro la Spagna , una	squadra di riserve uscite	sconfitte ma a testa alta . " Tutte le ragazze	Sport	
" les Biarrots " . In ogni caso , la	squadra che uscirà	trionfante dalla battaglia , avrà il morale alle stelle e	Sport	
il ritorno si disputerà allo stadio di Gubbio . Le	squadre che usciranno	sconfitte da questi confronti non saranno eliminate . Si sfideranno	Sport	
, segnare un gran bel gol . Se una grande	squadra vuole uscire	da una crisi pazzesca , deve vincere con il Chelsea	Sport	
, ma lo spettacolo è stato assicurato ed entrambe le	squadre sono uscite	dal Hüseyin Avni [...] Le età di Lulic	Sport	
riesce , ma nessuno si tira indietro » . La	squadra sa come uscire	da questo momento ? « Fino a un mese fa	Sport	
guarda avanti : " Adesso dobbiamo solamente cercare insieme alla	squadra di uscire	da un momento delicato guardando con ottimismo al futuro e	Sport	
momento positivo di Pato : « Il Milan è una	squadra che è uscita	fuori da un periodo di incidenti anche grazie all' ampia	Sport	
, giocato ieri sera tra Lazio e Reggina , la	squadra calabrese è uscita	sconfitta , rendendo impossibile il recupero sul quartultimo posto ,	Sport	
In soiaonia ad aprile : a Ostia è estate La	squadra di Mihailovic esce	incredibilmente dalla Coppa Italia : dono essere andata in vantaggio	Sport	

5. Patterns

Un pattern è una sequenza definita dall'utente di forme e/o lemmi e/o elementi appartenenti ad una pos e/o tokens, che può essere cercata nei corpora.

Come nel caso delle concordanze la ricerca si può indirizzare su tutto il corpus o su un solo sotto-corpus o su collezioni di sotto-corpora . Di default la ricerca è sull'intero corpus ridire.

I risultati ottenuti con la funzione pattern, diversamente dalle concordanze, non danno la possibilità di calcolo delle collocate e, per economia, possono essere ordinati solo per raggruppamento destro secondo lo stesso criterio illustrato in precedenza.

L'interfaccia aiuta l'utente a scrivere in modo semplice il pattern ovvero la sequenza da cercare nel corpus selezionato. Il suo utilizzo prevede una conoscenza delle relazioni grammaticali fondamentali e del modo in cui queste si manifestano in italiano.

L'utente ha la possibilità di scrivere sequenze comunque lunghe, ma ovviamente, si consiglia di definire sequenze di non più di 5 elementi.

Nella definizione di ciascun elemento della sequenza un menù a tendina permette la selezione delle 4 alternative disponibili (forma, pos o pos esteso, lemma), una volta riempito il primo campo si può aggiungere il secondo cliccando sul "+"

Il sistema propone sempre la possibilità di definire una distanza tra i parametri definiti (distanza 0, ovvero seguire immediatamente, è il default).

L'esistenza di una serie di elementi tra i due elementi definiti può essere opzionale o necessaria

Il seguente pattern definisce la sequenza in cui il lemma andare è seguito immediatamente dal lemma montagna

Pattern

1. Lemma 0 2. Lemma

Tutti i domini: Sì No

Pattern 1

Il seguente pattern definisce la sequenza tra andare e montagna con in mezzo da 0 a due tokens, e quindi prende sia “andare montagna” “andare in montagna” “andare per le montagne” “andare sulle belle montagne”.

Pattern

1. Lemma 2 2. Lemma

Tutti i domini: Sì No

Pattern 2

Il seguente pattern seleziona i contesti in cui andare e montagna sono necessariamente divisi da 2 tokens, ovvero esclude sia “andare montagna” sia “andare in montagna” e prende sia “andare sulla bellissima montagna” sia “andare lentamente in montagna”.

Pattern

1. Lemma 2 2. Lemma

Tutti i domini: Sì No

Pattern 3

La seguente è la concordanza ottenuta con il pattern 2

Risultati: 1 - 20 di 202

Sx		Dx	URL
. 42Altri discendenti di Simeone , circa cinquecento persone ,	andarono invece sulle montagne	di Seir , guidati da Pelatia , Nearia , Refaia	
reclama emozioni , passione , scoperte . Come quando si	va in montagna	, spesso un colle nasconde un' altra salita , un'	
Kobe ... Dalla Parola del giorno " Gesù se ne	andò sulla montagna	a pregare e passò la notte in orazione . Quando	
questo Paese , ogni bambino ha il sacrosanto diritto di	andar sulle montagne	russe e può mangiare tutto lo zucchero filato che vuole	
tutti dolcissimi e desiderosi di comprenderci . Mi piace molto	andare in montagna	con il cane (è un grande piacere !)	
. 12Il popolo di Moab si consumerà a furia di	andare sulle montagne	a pregare nei santuari e nei templi : non gli	
dimenticati (e come potrebbero ?) mais tout s'en	va , dalle montagne	di spazzatura gomorrana spunta un nuovo eroe del rione sanità	
la loro attenzione al cibo , mentre le strade che	vanno in montagna	attraversano paesaggi incontaminati e spettacolari . A Modena il Palazzo	
1970 , sempre a Padova . Ma non e '	andata meglio in montagna	. E qui il dato fa ancora piu ' scalpore	
) . Prima della elezione degli apostoli " se ne	andò sulla montagna	a pregare e passò la notte in orazione " (
bambino al mare vestito di tutto punto come se dovesse	andare in montagna	. Infine l' ultima che vive in uno stato di	
della velocità della luce , e se Maometto non può	andare alla montagna	, sarà la montagna ad andare da lui . Alla	
` cosi ` almeno diventa piu ` simpatica : sono	andato in montagna	questa estate , ho comprato una camicetta tirolese a mia	
Era una donna oriunda di Oliena ; una lavoratrice che	andava sulle montagne	per raccogliere scorza di ontano , con la quale faceva	
È per questo che non lo accendo mai , quando	vado in montagna	. Me lo porto , ma non lo accendo .	
È per questo che non lo accendo mai , quando	vado in montagna	. Me lo porto , ma non lo accendo .	
più grandi era che dopo tanti anni non era potuto	andare in montagna	, e pensare a quella casa vuota lo faceva precipitare	
al pilota . Dalla Parola del giorno Gesù se ne	andò sulla montagna	a pregare e passò la notte in orazione . Quando	

Attraverso la funzione di raggruppamento emergono in modo impressionante i modi in cui le due espressioni sono collegate di preferenza e le alternative possibili. La figura evidenzia i tipi più frequenti (marcati in azzurro), ma anche le associazioni avverbiali.

Raggruppamento a destra

Si/No N. forme: 0 Raggruppa per lemma forma

Cerca

« Prima pagina » « Precedente » **Prossima »** **Ultima pagina »**

Risultati: 1 - 20 di 30

Num.	Sx		Dx	Dominio	URL
+ 89 44,06%	anzi ti invita quasi a occupare il primo posto !	Andando in montagna	, là dove finiscono gli alberi e spunta solo la	Religione	
+ 40 19,8%	questo Paese , ogni bambino ha il sacrosanto diritto di	andar sulle montagne	russe e può mangiare tutto lo zucchero filato che vuole	Cinema	
+ 15 7,43%	scese a valle per il bene della pianura . «	Andate alla montagna	- diceva tempo fa una vignetta satirica raffigurando una frana	Informazione	
+ 8 3,98%	conditi con il tradizionale burro e formaggio del luogo .	Andando verso le montagne	della Valtellina merita una sosta la Latteria sociale di Delebio	Informazione	
+ 4 1,98%	... deve essere presente nelle montagne . È	andata nelle montagne	e in queste montagne del suo paese di Giudea ,	Religione	
+ 4 1,98%	Dove vai ? " gli chiedono , Risponde : "	Vado dalla montagna	" . " Perché ? " " Per spostarla .	Letteratura_e_Teatro	
+ 3 1,49%	addio !! fanciulli si dividono dai parenti , e	vanno allontanandosi sulla montagna	: di là si volgono , melanconici , stendono le	Musica	
+ 3 1,49%	dimenticati (e come potrebbero ?) mais tout s'en	va , dalle montagne	di spazzatura gomorrana spunta un nuovo eroe del rione sanità	Cinema	
+ 3 1,49%	anziché andare a combattere per la Repubblica di Salò sono	andati tra le montagne	. Lo spettacolo racconta del partigiano Giovanni Molinari (il	Letteratura_e_Teatro	
+ 2 0,99%	, Bella Stella , rendi noto il mio destino ,	vai oltre quelle montagne	, là dove si trova il mio cuoricino . Ma	Musica	
+ 2 0,99%	, la differenza la faranno le cronometro tra gente che	va forte in montagna	. Per questo bisognerà prestare attenzione anche a Gonchar .	Sport	
+ 2 0,99%	1970 , sempre a Padova . Ma non e '	andata meglio in montagna	. E qui il dato fa ancora piu ' scalpore	Informazione	
+ 2 0,99%	di fede : un percorso in salita , proprio come	andare verso una montagna	alta alta . Noi siamo in cammino dietro Gesù ,	Religione	

Le ricerche su pattern sono in funzione dell'interesse dell'utente. Per esempio se ci si chiede specificamente quali sono le preposizioni che legano "andare" e "montagna" a secondo della diversa selezione lessicale, il pattern sarà più definito:

Pattern

1. Lemma 1 2. PoS 1 3. Lemma

Tutti i domini: Sì No

I raggruppamenti del contesto ottenuto focalizzeranno quindi maggiormente su questa specifica associazione.

Raggruppamento a destra
 Sì/No N. forme: Raggruppa per lemma forma

Cerca

« Prima pagina | « Precedente | **Prossima** | » Ultima pagina »

Risultati: 1 - 20 di 40

Num.	Sx		Dx	Dominio	URL
+ 89 42,79%	anzi ti invita quasi a occupare il primo posto !	Andando in montagna	, là dove finiscono gli alberi e spunta solo la	Religione	
+ 40 19,23%	questo Paese , ogni bambino ha il sacrosanto diritto di	andar sulle montagne	russe e può mangiare tutto lo zucchero filato che vuole	Cinema	
+ 15 7,21%	scese a valle per il bene della pianura . «	Andate alla montagna	- diceva tempo fa una vignetta satirica raffigurando una frana	Informazione	
+ 8 3,85%	conditi con il tradizionale burro e formaggio del luogo .	Andando verso le montagne	della Valtellina merita una sosta la Latteria sociale di Delebio	Informazione	
+ 4 1,92%	... deve essere presente nelle montagne . È	andata nelle montagne	e in queste montagne del suo paese di Giudèa ,	Religione	
+ 4 1,92%	Dove vai ? " gli chiedono . Risponde : "	Vado dalla montagna	" . " Perché ? " " Per spostarla .	Letteratura_e_Teatro	
+ 3 1,44%	addio ! I fanciulli si dividono dai parenti , e	vanno allontanandosi sulla montagna	: di là si volgono , melanconici , stendono le	Musica	
+ 3 1,44%	dimenticati (e come potrebbero ?) mais tout s'en	va , dalle montagne	di spazzatura gomorrjana spunta un nuovo eroe del rione sanità	Cinema	
+ 3 1,44%	anziché andare a combattere per la Repubblica di Salò sono	andati tra le montagne	. Lo spettacolo racconta del partigiano Giovanni Molinari (il	Letteratura_e_Teatro	
+ 2 0,96%	, Bella Stella , rendi noto il mio destino ,	vai oltre quelle montagne	, là dove si trova il mio cuoricino . Ma	Musica	
+ 2 0,96%	sarà Lui stesso ad occuparsi del proprio gregge perché non	vada disperso fra le montagne	e non sia preda dei lupi rapaci . Gli fanno	Religione	
+ 2 0,96%	, la differenza la faranno le cronometro tra gente che	va forte in montagna	. Per questo bisognerà prestare attenzione anche a Gonchar .	Sport	
+ 2 0,96%	1970 , sempre a Padova . Ma non e '	andata meglio in montagna	. E qui il dato fa ancora piu ' scalpore	Informazione	

6. Gli *Sketches* delle parole

6.1 *Lemma* singolo

L'infrastruttura di ricerca di RIDIRE è implementata secondo le prospettive che si sono affermate nell'ultimo decennio per l'interrogazione di grandi corpora in rete e in particolare sviluppa le funzioni di *sketch*, che permettono di osservare in modo complessivo la serie di alternative dell'uso linguistico in relazione alle varie relazioni grammaticali della parola.² Gli *sketches* uniscono la possibilità di investigare le relazioni grammaticali delle espressioni, propria dei *pattern* di ricerca con l'informazione delle collocate e sono definiti come il *sommario* delle relazioni grammaticali e delle collocazioni di una parola (KILGARIFF et al. 2004).

Con gli *sketches* viene evidenziato il diverso valore grammaticale delle preferenze associative di una parola. Piuttosto che fornire all'utente una unica lista di collocazioni, nelle quali tutte le relazioni tra testa e collocato sono unite insieme, si restringono le collocate alle sole significative in relazioni grammaticali determinate, come ad esempio la relazione soggetto-oggetto, complemento indiretto del nome o del verbo. Ne risulta l'insieme delle relazioni instaurate da una parola che può essere mostrato in un unico quadro riassuntivo o una relazione per volta.

² Si veda in particolare l'implementazione di questa funzione nell'omonima infrastruttura *Sketch Engine* che raccoglie un'ampia serie di grandi corpora interrogabili per molte lingue diverse.

Gli *sketches* sono calcolati in RIDIRE per tutti i lemmi nominali, verbali, aggettivali e avverbiali con frequenza minima di 500 occorrenze.

Le seguenti sono le relazioni di cui sono computabili gli *sketches*, che sono riportati nel menù a tendina evidenziato nella figura seguente, e possono essere selezionati tutti insieme o singolarmente scegliendo la sigla dello *sketch* desiderato.³

Relazioni Nome-Verbo

postV_N: dato un verbo restituisce i nomi che lo seguono (oggetti diretti o indiretti)

preV_N : dato un verbo restituisce i nomi che lo precedono (i soggetti)

preN_V : dato un nome restituisce i verbi che lo precedono (oggetto o oggetto indiretto)

postN_V : dato un nome restituisce i verbi che lo seguono (di cui è soggetto)

Legami preposizionali instaurati da un lemma

prep-%s : dato un lemma (nome, verbo, aggettivo) restituisce le modificazioni preposizionali che instaura

Relazioni di modificazione tra nomi

n_modifies: dato un nome restituisce i nomi che lo precedono, direttamente o separati da aggettivo o nome

n-modifier: dato un nome restituisce i nomi che lo seguono, direttamente o separati da aggettivo o nome (modificatori del nome)

Relazione Nome Aggettivo

NdiA: dato un nome restituisce gli aggettivi che lo seguono

AdiN: dato un aggettivo restituisce i nomi che lo precedono

Relazione Verbo Avverbio

postV_ADV : dato un verbo restituisce gli avverbi che lo seguono

preV_ADV : dato un verbo restituisce gli avverbi che lo precedono

Selezioni preferenziali nella congiunzione tra lemmi

e_o : dato un lemma di qualsiasi PoS restituisce i lemmi di uguale PoS che collocano con questo in relazione di congiunzione o disgiunzione.

³ La definizione formale delle ricerche per la ricerca degli *sketches* non è riportata qui. Nell'infrastruttura computazionale del test corpus di RIDIRE è stata utilizzata la stessa serie di espressioni regolari utilizzata in *Sketch Engine*, proposta da Marco Baroni ed elaborata da Valentina Efrati e Francesca Masini.

Anche in questo caso, è possibile interrogare l'intero corpus o alternativamente i vari sotto-corpora, a seconda dell'interesse dell'utente. Ovviamente è quest'ultimo il valore aggiunto specifico di RIDIRE. L'utente può selezionare le opzioni da un menù a tendina.

Selezione delle opzioni per dominio e tipo di *sketch*

Gli esempi seguenti evidenziano due degli usi più rilevanti degli *sketches*, che possono riguardare rispettivamente domande relative all'intero corpus, e quindi alla lingua italiana in generale o relative alle differenze d'uso tra un dominio e l'altro.

La Figura seguente, relativa ad una *query* sull'intero corpus, evidenzia l'insieme delle associazioni preferenziali che il verbo uscire trova nelle sue principali relazioni grammaticali (avverbio-verbo, soggetto-verbo, verbo-oggetto, verbo-preposizione).

preADV_V	32665924	preV_N	77875906	postV_N	38480374	e_o	50411977	
presto	48902	5,97	sipario	3631	5,98	abbassare	11408	10,21
ulteriormente	20322	5,01	braccio	30796	5,37	sorridere	16286	5,85
già	2090	4,74	mattino	14853	5,24	gridare	16354	5,74
lentamente	9172	4,66	mattina	51666	4,94	alzare	28535	5,04
notevolmente	7221	4,4	fanciulla	4446	4,85	guardare	141937	4,47
leggermente	10531	4,39	fumo	10246	4,71	camminare	21411	4,39
troppo	130571	4,23	dito	13954	4,63	correre	53769	4,06
bruscamente	1790	4,2	pollice	2270	4,41	bagnare	3760	4,02
nuovamente	21584	4,14	pugno	8436	4,22	piangere	19445	3,99
tardi	12894	3,83	sentinella	1307	4,13	passaggiare	5568	3,85
gradualmente	4518	3,31	popolo	101088	4,08	urlare	8767	3,8
nemmeno	37702	3,31	città	10158	3,85	allargare	18465	3,77
improvvisamente	8288	3,29	lupo	8975	3,81	stendere	9639	3,56
intorno	63750	3,21	vento	29734	3,69	mangiare	29045	3,54
subito	177642	3,2	ruota	11234	3,53	uscire	117871	3,42
verticalmente	1066	3,15	nube	5904	3,51	abbandonare	43016	3,2
definitivamente	17056	3,11	opposizione	39992	3,46	tagliare	24962	3,17
ancora	471411	3,07	broker	1986	3,42	percuotere	1195	3,09
sopra	108194	3,07	gomito	2292	3,41	tirare	31266	3,07
immediatamente	38261	3,06	re	50299	3,38	applaudire	4396	2,99

pp_sul	2022713		pp_di	9900498	
pedale	1206	4,53	scatto	11245	6,55
corpo	123366	2,37	mattino	14853	5,5
traguardo	15965	1,88	botto	2494	3,25
viso	16352	1,87	centimetro	7204	2,87
gomito	2292	1,67	tiro	15240	2,75
palco	16230	0,55	tono	29400	2,64
fianco	26441	0,25	notte	96022	2,53
predellino	232	0,19	peso	50286	2,43
piegino	465	0,18	sopracciglio	1195	1,73
malaffare	548	0,17	colpo	52775	1,69
mare	59267	0,16	quarto	83401	1,55
reazionario	1789	0,11	testa	72840	1,54
corner	2292	0,09	decina	23944	1,32
scoglio	2652	0,07	decibel	515	1,17
scorcio	2961	0,06	ottava	1754	1,11
canottaggio	3029	0,05	metro	78418	1,1
pullman	3498	0,03	voce	153236	0,85
pelo	4207	0	livello	245454	0,84
lido	5884	-0,07	braccio	30796	0,73
gradino	6286	-0,09	centesimo	12552	0,67

pp_a	8081761		pp_da	4353990	
fatica	25124	2,46	tavola	45750	5,16
mezzogiorno	4735	2,3	terra	127855	4,25
campanile	1749	2,11	letto	30464	3,54
stento	2793	2,06	sedia	12042	2,5
metà	17487	1,83	gendarme	410	1,76
barriera	15490	1,57	bidonville	249	1,19
miliardo	94618	1,55	collinetta	318	1,18
dispetto	5481	1,53	moralista	1298	1,14
prua	1059	1,15	sud-est	1512	1,13
braccio	30796	1,14	tappeto	7269	0,87
sufficienza	2418	1,08	equipaggio	10523	0,75
spirale	3568	1,03	sole	68482	0,43
centrocampo	6802	0,89	acquitrino	116	0,19
altezza	24099	0,32	crocchio	224	0,19
tratto	54435	0,24	cespuglio	1206	0,14
piede	56649	0,21	pigrizia	1348	0,13
tagliola	165	0,19	striscia	5075	-0,04
grappolo	1165	0,14	vasca	6820	-0,11
dismisura	1176	0,14	poltrona	7241	-0,13
volo	31497	0,13	platea	8355	-0,17

Lo sketch delle principali relazioni di “uscire”

La Figura seguente corrisponde alla comparazione dei risultati ottenuti calcolando gli sketch singolarmente degli oggetti preferenziali del verbo “alzare” nei domini *Cucina* e *Religione*. Tale comparazione può essere resa standard

Sketch

Lemma: PoS:

Scegli sketch:

Scegli un dominio:

Calcola sketch

postV_N	9599
fiamma	432
bollore	269
bordo	510
fuoco	2759
gomito	7
bandiera	12
gas	28
lembo	30
sguardo	49
mattino	70
mattina	121
occhio	208
temperatura	619
mano	865
cibo	639
sapore	657
forno	2324

Sketch

Lemma: PoS:

Scegli sketch:

Scegli un dominio:

Calcola sketch

postV_N	911115
occhio	17436
sguardo	10210
voce	18143
mano	23314
grido	3614
tono	959
calice	1713
spada	2695
testa	3002
mattina	2146
capo	11826
lamento	995
barriera	941
tiro	151
braccio	3494
velo	1130
bandiera	251

Sketches degli "oggetti" del verbo "alzare" nei domini Cucina e Religione

6.2 Differenze tra parole e differenze di una parola nei domini

RIDIRE propone due ulteriori metodi per sfruttare la metodologia degli Sketches.

La prima (sketch difference) già utilizzata in Sketch Engine, è un metodo per confrontare sinteticamente, attraverso le collocazioni, le differenze d'uso di due parole (Nomi, Verbi, Aggettivi o Avverbi), di solito due parole simili delle quali si vuol conoscere la diversa sfera associativa, in tutto il corpus o in uno specifico dominio.

RIDIRE ▲

- [Entra](#)

Querying Area ▲

- [Liste di frequenza](#)
- [Concordanze](#)
- [Pattern](#)

Sketch Area ▲

- [Lemma singolo](#)
- [Confronto due lemmi](#)
- [Confronto due domini](#)

Differenza sketch tra due lemmi

Primo lemma: PoS: nome ▼

Secondo lemma:

[Mostra differenze](#)

Dominio

- Tutti
- Architettura e design
- Arti figurative
- Cinema
- Cucina
- Letteratura e Teatro
- Moda
- Musica
- Religione
- Sport

- Amministrazione e Legislazione
- Economia e Affari
- Informazione

La seconda modalità di interrogazione invece, specificamente realizzata in RIDIRE per meglio sfruttare la sua divisione in domini d'uso, consente la comparazione dell'uso di una parola in due domini diversi.

RIDIRE ▲

- [Entra](#)

Querying Area ▲

- [Liste di frequenza](#)
- [Concordanze](#)
- [Pattern](#)

Sketch Area ▲

- [Lemma singolo](#)
- [Confronto due lemmi](#)
- [Confronto due domini](#)

Differenza sketch tra due domini

Lemma: PoS: **nome** ▼

Primo dominio

- Tutti
- Architettura e design
- Arti figurative
- Cinema
- Cucina
- Letteratura e Teatro
- Moda
- Musica
- Religione
- Sport
- Amministrazione e Legislazione
- Economia e Affari
- Informazione

Secondo dominio

- Tutti
- Architettura e design
- Arti figurative
- Cinema
- Cucina
- Letteratura e Teatro
- Moda
- Musica
- Religione
- Sport
- Amministrazione e Legislazione
- Economia e Affari
- Informazione

Mostra differenze

Entrambi i tipi di Sketches forniscono le differenze, rispettivamente del lemma o di due lemmi in tutti gli Sketches disponibili.

Sono evidenziate in verde le associazioni principali del primo lemma e in rosso quelle del secondo. I colori di evidenziazione sfumano quando il contrasto è più ridotto. Sono in bianco le associazioni comuni ad entrambi i lemmi.

Nella Figura seguente possiamo osservare il risultato della ricerca relativa alla differenza tra i due aggettivi italiani, *forte* e *resistente*, nel dominio della *moda*. Evidenziamo in particolare due degli sketches tra i più rilevanti per l'aggettivo nei quali i due aggettivi sono stati comparati: nella prima schermata *e_o*, che identifica le associazioni in cui l'aggettivo è in opposizione o in coordinazione con un altro, nella seconda schermata lo sketch *NofA*, che evidenzia i nomi modificati di preferenza da questi aggettivi.

	NofA				e_o				
impatto	365	0	10,52	0	deciso	159	0	10,38	0
personalità	291	0	10,02	0	chiaro	59	0	8,31	0
legame	137	0	9,21	0	sicuro	38	0	8,14	0
tinta	169	0	9,18	0	indipendente	23	0	8,05	0
pezzo	254	0	9,15	0	determinato	17	0	7,93	0
carattere	122	0	8,87	0	sensuale	38	0	7,91	0
crescita	174	0	8,85	0	riconoscibile	16	0	7,7	0
identità	108	0	8,85	0	grintoso	19	0	7,67	0
appeal	73	0	8,36	0	sano	18	0	7,65	0
espansione	79	0	8,35	0	coraggioso	12	0	7,46	0
richiamo	71	0	8,26	0	emerso	10	0	7,44	0
presenza	99	0	8,24	0	vivace	19	0	7,37	0
colore	374	0	8,19	0	pratico	0	13	0	7,29
segnale	63	0	8,16	0	aerodinamico	0	3	0	7,31
emozione	70	0	8,12	0	protettivo	0	6	0	7,34
contrasto	94	0	8,09	0	funzionale	0	10	0	7,6
punto	152	0	7,96	0	duttile	0	4	0	7,76
messaggio	51	0	7,67	0	elastico	0	10	0	7,82
connotazione	39	0	7,61	0	leggero	0	56	0	7,94
impronta	40	0	7,57	0	comodo	0	22	0	7,94
contenuto	49	0	7,53	0	robusto	0	9	0	8,54
carica	38	0	7,43	0	capiente	0	13	0	8,63
polimero	0	2	0	7,73	impermeabile	0	24	0	8,96
guaina	0	3	0	8,14	flessibile	0	19	0	9,44
ultra	0	4	0	9,24	ultra	0	11	0	9,51

forte
-6.0
-4.0
-2.0
0.0
2.0
4.0
6.0
resistente

Sketch Difference di *forte* e *resistente* nel dominio Moda

Dall'esempio si evidenzia che *forte* ha usi più variati nella moda ed è spesso legato a tratti di personalità, mentre *resistente* è un requisito degli oggetti e dei tessuti.

La funzione di Sketch difference, in Ridire può mostrare anche il contrasto tra le associazioni che una parola può trovare in due domini diversi. Nella ricerca seguente si evidenzia la differenza d'uso dell'aggettivo *forte* rispettivamente nei domini Cucina e Religione.

		NofA				e_o				
	farina	1034	0	12,44	0	penetrante	20	0	9,27	0
	abbraccio	63	0	8,95	0	rispettoso	22	0	9,26	0
	odore	75	0	8,64	0	acidulo	16	0	8,44	0
	sapore	259	0	8,3	0	forte	70	0	8,19	0
	calore	49	0	7,84	0	tedesco	25	0	8,16	0
	influenza	26	0	7,73	0	marcato	11	0	8,13	0
	tinta	23	0	7,72	0	sgradevole	10	0	7,89	0
	emicrania	16	0	7,27	0	profumato	21	0	7,48	0
	lama	17	0	7,22	0	senape	11	0	7,33	0
	gusto	90	0	7,12	0	aromatico	18	0	7,18	0
	espressione	22	71	7,45	6,6	maturo	31	21	8,48	6,68
	emozione	22	36	7	6,97	debole	12	29	7,4	6,6
	legame	22	116	7,51	8,19	deciso	20	67	8,61	8,59
	vento	15	83	6,94	8,05	sano	11	63	7,02	7,97
	contrasto	11	62	6,37	7,64	coerente	0	28	0	7,16
	personalità	12	88	6,73	8,08	saldo	0	34	0	7,27
	stimolo	0	42	0	7,17	sereno	0	47	0	7,53
	appello	0	66	0	7,38	soave	0	31	0	7,53
	esperienza	0	159	0	7,4	robusto	0	31	0	7,7
	invito	0	100	0	7,45	tenero	0	40	0	7,79
	impulso	0	56	0	7,5	generoso	0	99	0	7,8
	carattere	0	86	0	7,59	chiaro	0	110	0	8,03
	tensione	0	70	0	7,78	dolce	0	61	0	8,14
	grido	0	120	0	8,42	potente	0	94	0	8,22
	richiamo	0	215	0	9,34	coraggioso	0	138	0	9,17

Cucina	
-6.0	
-4.0	
-2.0	
0.0	
2.0	
4.0	
6.0	
Religione	

Sketches Domain: *forte* (Cucina vs. Religione)

Si evidenzia facilmente che *forte* è applicator a sapori nel dominio della cucina e a sentimenti nel dominio della religione.