

Sviluppo e test di tecniche di Natural Language Processing per l'Analisi Reputazionale

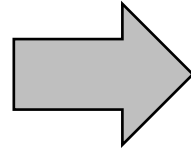
Relatore: *Prof.ssa Elisabetta Fersini*
Co-relatore: *Dott. Andrea Catacchio*

Relazione della prova finale di:
Lorenzo Lorgna
Matricola 829776

Riciclaggio e Analisi reputazionale

1

Il reato di **riciclaggio** consiste nell'investire capitali provenienti da reato all'interno di attività lecite in modo tale da ripulire tali somme di denaro



L'**analisi reputazionale** rappresenta uno dei cardini di queste attività di contrasto e prevenzione

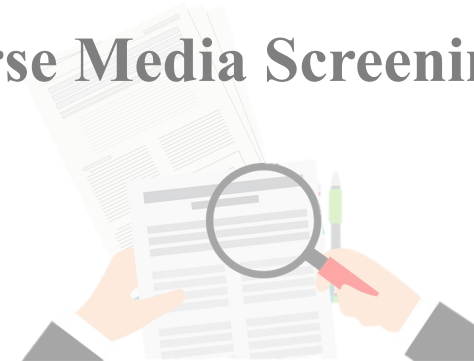


590 - 3200 miliardi

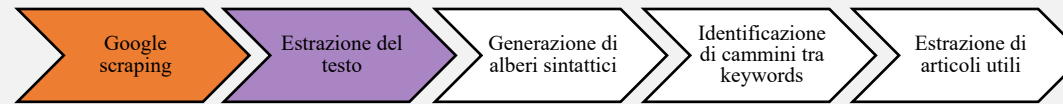


Ricerche sul **web** basate su specifiche keywords di **articoli** che sono relativi al soggetto cercato

Adverse Media Screening







I Per ogni soggetto vengono effettuate tre ricerche:

Nome + Cognome +

Condannato

Condannata

Misure cautelari



II Per le prime tre pagine risultanti da ciascuna ricerca vengono memorizzati:

- Titolo degli articoli
- URL degli articoli
- Data di pubblicazione degli articoli
- Data di scaricamento degli articoli
- Indice di pagina di Google

III

Pirella, il sindaco indagato per bancarotta

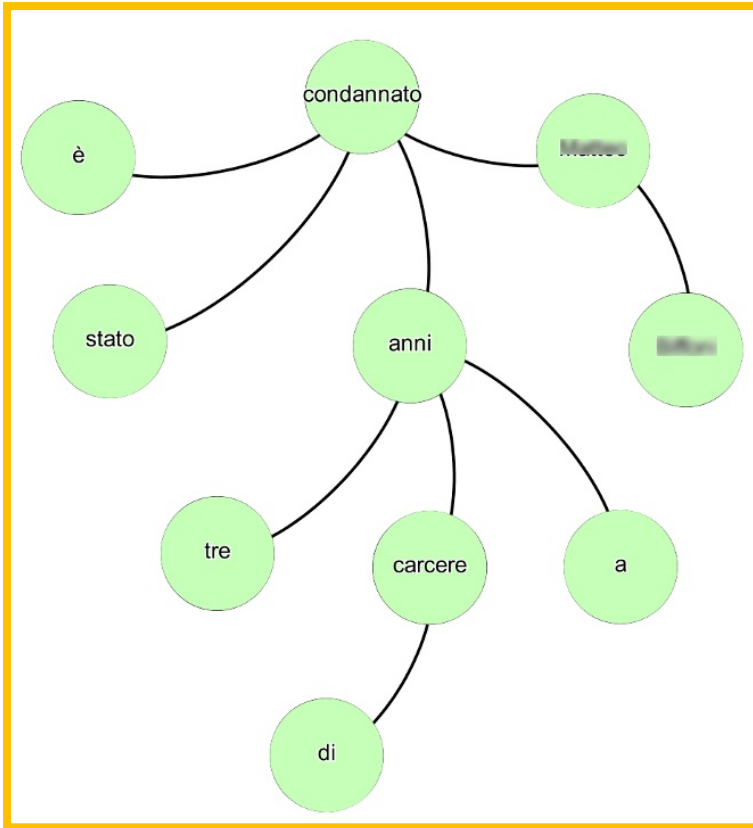
Chiamato a comparire dalla procura in quanto presidente della Provincia in un'indagine sulle sorti del centro di ricerca pubblico fallito



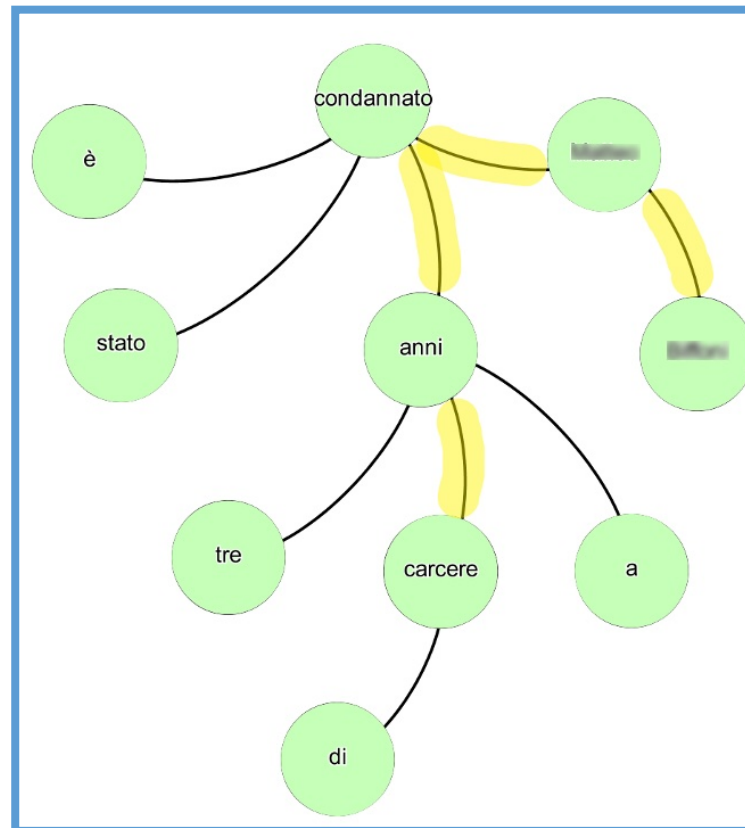
A partire dall'URL di ciascun articolo ne viene estratto il relativo testo



Generazione di alberi sintattici



Identificazione di cammini tra keywords



Estrazione di articoli utili

Articolo selezionato se:

- ✓ Cammino minimo tra nome e keyword < 6
- ✓ Cammino minimo tra cognome e keyword < 6
- ✓ Pattern Nome + Cognome

Esempio di Output:

Titolo: Fallimento del C***f, citato a giudizio M****o B*****i - gonews.it
Link: https://www.gonews.it/2018/09/18/fallimento-del-c***f-citato-a-giudizio-m****o-b*****i/
Data di scaricamento: 12 lug 2020
Data di pubblicazione: 18 set 2018
Frasi estratte: 'Con B*****i citati a giudizio per bancarotta semplice ex amministratori del C***f ed ex membri del cda, L**a R*****i, L***a C*****i, L*****o G****i, V*****a M****i, G*****o B***a, oltre ai tre componenti del collegio sindacale, M*****o L*****i, M*****o P****i e M***o B**i.' 'Paradossali le dichiarazioni del Sindaco B*****i, che anche di fronte al rinvio a giudizio sostiene di aver trovato in eredità dalle precedenti amministrazioni questa situazione.' 'Citazione a giudizio per M****o B*****i, sindaco di P***o.'

Analisi sperimentale

Esempio di analisi di un caso di studio:

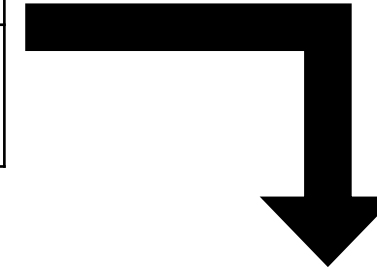
<i>Caso di studio</i>	<i>Dependency parser</i>	<i>Risultati ricerche Google</i>	<i>Articoli estratti</i>	<i>Articoli ritenuti utili</i>	<i>Tempo medio (min)</i>
D****o C*****a	spaCy	68	38	38	2,1
	Stanza	68	38	38	2,4



<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
0,823	1	0,760	0,863

Esempio di analisi di un caso di studio:

<i>Caso di studio</i>	<i>Dependency parser</i>	<i>Risultati ricerche Google</i>	<i>Articoli estratti</i>	<i>Articoli ritenuti utili</i>	<i>Tempo medio (min)</i>
Y*****g L*	spaCy	59	19	13	2
	Stanza	59	19	13	2,2



<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
0,694	0,684	0,520	0,590

Sistema basato su clausole

Named Entity Recognition

Frase attive e passive

Supporto multi lingua

Grazie per l'attenzione