



Il tuo partner per la Formazione e la Trasformazione digitale della tua azienda



Note			



### **SPRING AI**

### GENERATIVE ARTIFICIAL INTELLIGENCE CON JAVA

### Simone Scannapieco

Corso base per Venis S.p.A, Venezia, Italia

Settembre 2025

	LANGUAGE ARTIFICIAL INTELLIGENCE
Note	

# LANGUAGE AI ...o NLP



- n
- Sottocampo dell'Al dedicato allo sviluppo di tecnologie per il linguaggio umano
  - Comprensione
  - Elaborazione
  - Generazione
- Utilizzato intercambiabilmente con Natural Language Processing (NLP)
- ▲ Trasversale rispetto alla classificazione canonica

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		







ı	V	U.	t	6
		$\mathbf{\mathcal{C}}$	L	·

### LANGUAGE MODELS ... O MODELLI DEL LINGUAGGIO



- Modelli che simulano la comprensione e la generazione del linguaggio umano con approcci statistici e modelli della rete neurale
- Predicono la parola successiva in una sequenza in base al contesto. . .
- occidenta di un dizionario calcolando probabilità su ogni singola parola di un dizionario
- A Qualunque task NLP può essere trasformata in un problema di generazione di testo!
  - (thinking) Ho la frase «Mi piace come recita Hugh Laurie!», e devo determinarne il sentiment —
  - 🚨 "Il sentiment della frase «Mi piace come recita Hugh Laurie!» é: "
  - >\_ (thinking) Eseguo l'inferenza...—
  - >\_ (thinking) Ho ottenuto le seguenti probabilità come prossimo token da generare: «positivo» al 45%, «negativo» al 2%, «cane» al 0.5%, «gatto» al 0.3%,...—
  - >\_ "Il sentiment della frase «Mi piace come recita Hugh Laurie!» é: positivo"

Processo di sentiment analysis

🐸 Simone Scannapieco	Spring AI - Corso base	5/17

Note		

### LANGUAGE MODELS ... O MODELLI DEL LINGUAGGIO



- O Modelli che simulano la comprensione e la generazione del linguaggio umano con approcci statistici e modelli della rete neurale
- Predicono la parola successiva in una sequenza in base al contesto. . .
- calcolando probabilità su ogni singola parola di un dizionario
- A Qualunque task NLP può essere trasformata in un problema di generazione di testo!
  - (thinking) Ho la domanda «Chi ha scritto "L'origine della specie"?» e voglio ottenere la risposta —
  - "D: Chi ha scritto «L'origine della specie»? R: "
  - >\_ (thinking) Eseguo l'inferenza...—
  - >\_ (thinking) Ho ottenuto le seguenti probabilità come prossimo token da generare: «Charles» al 25%, «Darwin» al 15%, «cane» al 0.2%, «gatto» al 0.1%,...-
  - >\_ "D: Chi ha scritto «L'origine della specie»? R: Charles"
  - 📤 "D: Chi ha scritto «L'origine della specie»? R: Charles "

  - >\_ (thinking) Eseguo l'inferenza...— >\_ (thinking) Ho ottenuto le seguenti probabilità come prossimo token da generare: «Darwin» al 65%, «Charles» al 1%, «cane» al 0.2%, «gatto» al 0.1%,...-

>\_ "D: Chi ha scritto «L'origine della specie»? R: Charles Darwin"

Processo di question answering

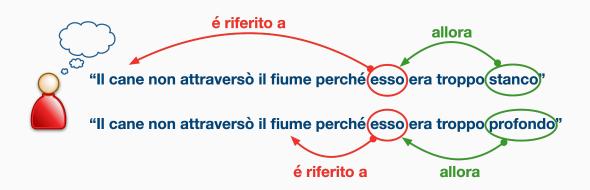
Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		





- 1 l significato di alcune parole è dipendente dal significato di altre (parole contestuali)
- ¿ L'umano affronta questo processo in maniera inconscia ed istantanea...
- ... ma una macchina?!

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note	

### LE ORIGINI: ELIZA (1966) IL PRIMO CHATBOT DELLA STORIA



- Sviluppato da Joseph Weizenbaum al MIT
- Simulava una conversazione con uno psicoterapeuta rogersiano
- Utilizzava semplici pattern matching e regole di sostituzione
- Dimostrava quanto facilmente le persone potessero essere ingannate da un programma semplice
- Primo esempio di illusione di comprensione da parte di una macchina

☑ Implementazione ELIZA

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		



8/17

Trasformazioni grammaticali: regole sintattiche applicate all'input utente

# Regole di trasformazione ELIZA ''I am'' -> ''you are'' ''my'' -> ''your'' ''me'' -> ''you'' ... "I am feeling sad today" >\_ (thinking) - you are feeling sad today ELIZA, elaborazione, 1966

m Venis S.p.A, Venezia, IT

Spring AI - Corso base

Simone Scannapieco

Note			





€ Template di risposta: frasi predefinite con slot per le sostituzioni

### Esempi di template ELIZA

```
''Tell me more about ___''
''What else comes to mind when ___?''
''Why ___?''
...
```

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

N	ot	e
---	----	---



10 / 17

m Venis S.p.A, Venezia, IT

Regole di pattern matching: riconoscimento di parole chiave nell'input e scelta fra possibili pattern correlati

# | ''I think about \_\_\_'' -> ''Tell me more about \_\_\_'' | ''I am thinking of \_\_\_'' -> ''What else comes to mind when you think of \_\_\_?'' | ''I am \_\_\_'' -> ''Why do you think \_\_\_?'' | ... | "I am feeling sad today" | ''Why do you think you are feeling sad today?" | ELIZA, esempio di risposta, 1966

Spring AI - Corso base

Simone Scannapieco

Note		



Strategia di fallback: formule generiche quando non trova pattern

"My whole world is falling apart."
>\_ "I see."
(oppure)
>\_ "Can you elaborate on that?"
(oppure)
>\_ "I see."
(oppure)
>\_ "What does that suggest to you?"
(oppure)
>\_ "Please go on."
(oppure)
>\_ "How does that make you feel?"

ELIZA, esempio di risposta, 1966

Simone Scannapieco

Spring AI - Corso base

🏛 Venis S.p.A, Venezia, IT

Note		

# BAG-OF-WORDS RAPPRESENTAZIONE TESTO NON STRUTTURATO



12/17

- Menzionato per la prima volta negli anni '50, popolare negli anni 2000
- Metodo per rappresentare il testo non strutturato in formato numerico
- 1 linguaggio è complicato per i calcolatori
  - Dil testo perde significato quando rappresentato da 0 e 1

Spring AI - Corso base

Simone Scannapieco

➡ Focus principale: rappresentare il linguaggio in modo strutturato per l'uso da parte dei calcolatori

m Venis S.p.A, Venezia, IT

Note	

### **BAG-OF-WORDS: COME FUNZIONA** PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO



- 1 Tokenizzazione: processo di divisione delle frasi in parole individuali o sotto-parole (token) Rispetto a un delimitatore (specifico o wildcard) e una blacklist
- Creazione del vocabolario: estrazione di entità uniche dai token
  - Applicando estrazione della radice (stemming)
  - Vettore ordinato di radici
- Conteggio delle parole: rappresentazione numerica basata sulla frequenza



ote	

### BAG-OF-WORDS: COME FUNZIONA PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO



Creazione del vocabolario

m Venis S.p.A, Venezia, IT

13 / 17

Note \_\_\_\_\_\_

Spring AI - Corso base

Simone Scannapieco

### BAG-OF-WORDS: COME FUNZIONA PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO



- 1 Tokenizzazione: processo di divisione delle frasi in parole individuali o sotto-parole (token)
  - Rispetto a un delimitatore (specifico o wildcard) e una blacklist
- 2 Creazione del vocabolario: estrazione di entità uniche dai token
  - Applicando estrazione della radice (stemming)
  - ▲ Vettore ordinato di radici
- 3 Conteggio delle parole: rappresentazione numerica basata sulla frequenza

					V	ocab	ola	rio								
	che	bell-	can-	avere	ma	il		mi-	1	gatt-	I	essere	1	più	Ī	bell-
$\frac{1}{2}$	2	1 0	1 0	1 0	0 1	0 1		0 1		0 1		0 1		0 1		0

<b>l</b> "Che ∟ bel ∟ can	e che _ hai!"	
<u> </u>	mio _gattoè _più	bello"

Vettori bag-of-words

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note	

# BAG-OF-WORDS



- Perdita dell'ordine: frasi diammetralmente opposte hanno la medesima rappresentazione
- Mancanza di semantica: Non cattura il significato delle parole
- Nessuna generalizzazione: i sinonimi sono trattati come elementi totalmente separati
- Alta dimensionalità: Vocabolari enormi con molti zeri

	V	oc	abolario		
	Simone	1	mangiare	Į-	insalat-
$\frac{1}{2}$	1 1		1 1	1 1	1 1



Problema dell'ordine delle parole

Simone Scannapieco

Spring AI - Corso base

🏛 Venis S.p.A, Venezia, IT

Note		

### BAG-OF-WORDS LIMITAZIONI



- Perdita dell'ordine: frasi diammetralmente opposte hanno la medesima rappresentazione
- Mancanza di semantica: Non cattura il significato delle parole
- Nessuna generalizzazione: i sinonimi sono trattati come elementi totalmente separati
- Alta dimensionalità: Vocabolari enormi con molti zeri
  - 4 "Concetti come «Re» e «Regina» dovrebbero essere correlati..."
  - "Sinonimi come «felice» e «gioioso» dovrebbero avere una rappresentazione simile..."
  - Le con parole come «pitone» e «serpente», una generalizzazione dell'altra?!"

Mancanza di relazioni semantiche

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note			

### BAG-OF-WORDS LIMITAZIONI



- Perdita dell'ordine: frasi diammetralmente opposte hanno la medesima rappresentazione
- Mancanza di semantica: Non cattura il significato delle parole
- Nessuna generalizzazione: i sinonimi sono trattati come elementi totalmente separati
- Alta dimensionalità: Vocabolari enormi con molti zeri

♣ "Fornisci una stima del vocabolario usato in Wikipedia." > \_ "[...] English Wikipedia: ordine di grandezza  $10^5 - 10^7$  [...]. Tutte le Wikipedie insieme (tutte le lingue, forme di parola): ordine di grandezza  $10^5 - 10^8$  [...].

"Quindi un documento Wikipedia in Bag-Of-Words sarebbe rappresentato da un vettore di dimensione almeno 10<sup>5</sup>?!"

Problema della dimensionalità

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note			

### WORD EMBEDDINGS (2013) WORD 2 VEC E LE RAPPRESENTAZIONI DENSE



15/17

m Venis S.p.A, Venezia, IT

- Word2Vec: primo tentativo di successo per catturare il significato del testo negli embeddings
- Embeddings: rappresentazioni vettoriali di dati che tentano di catturarne il significato
- Addestrato su enormi quantità di dati testuali
  - British National Corpus
  - English Wikipedia
  - Google News
  - English Gigaword
  - **()** ...

👺 Simone Scannapieco

- Utilizza reti neurali per generare rappresentazioni semantiche
- ▲ Grandezza della rappresentazione vettoriale limitata a priori

Spring AI - Corso base

Note

### WORD2VEC: COME FUNZIONA APPRENDIMENTO DELLE RELAZIONI TRA PAROLE



- Principio fondamentale: parole che appaiono in contesti simili tendono ad avere significati simili
- Addestramento: predire se due parole sono vicine in una frase
- Risultato: parole con significati simili hanno embeddings vicini nello spazio

🖺 "[...] Il mio cane ama dormire nella sua cuccia [...]"

[...] Il veterinario ha deciso di sterilizzare il gatto [...]

\*[...] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario [...]"

\*[...] Ho comprato una cuccia per il mio gatto, ma continua a preferire il divano! [...]"

>\_ (thinking) — Devo capire quali termini condividono gli stessi contesti linguistici...—

Principio alla base di Word2Vec

You shall know a word by the company it keeps.

Ipotesi distributiva, John Rupert Firth, 1957

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		

### WORD2VEC: COME FUNZIONA APPRENDIMENTO DELLE RELAZIONI TRA PAROLE



- Principio fondamentale: parole che appaiono in contesti simili tendono ad avere significati simili
- Addestramento: predire se due parole sono vicine in una frase
- Risultato: parole con significati simili hanno embeddings vicini nello spazio

1 "[...] Il mio cane ama dormire nella sua cuccia [...]"

"[...] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario [...]"

\*[...] Ho comprato una cuccia per il mio gatto, ma continua a preferire il divano! [...]"

>\_ (thinking) — Devo avvicinare le rappresentazioni di «cane» e «gatto» nello spazio perché condividono spesso gli stessi vicini «veterinario» e «cuccia»! —

Addestramento di Word2Vec

Simone Scannapieco

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		

### WORD2VEC: COME FUNZIONA APPRENDIMENTO DELLE RELAZIONI TRA PAROLE



- Principio fondamentale: parole che appaiono in contesti simili tendono ad avere significati simili
- Addestramento: predire se due parole sono vicine in una frase
- Risultato: parole con significati simili hanno embeddings vicini nello spazio
  - >\_ "«Computer» è simile a «laptop», «pc», «desktop», «workstation», ..."
  - >\_ "«Re» sta a «uomo» come «Regina» sta a «donna»!"
  - >\_ "«Lunedì», «martedì», «mercoledì», ...formano un unico gruppo."

Proprietà emergenti: sinonimie, classificazioni, relazionalità

Portale Word2Vec

Simone Scannapieco Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		

### WORD2VEC



- Polisemia: confusione con parole utilizzabili con diversi significati in contesti diversi
- Contesti insufficienti: parole rare potrebbero non avere abbastanza esempi

>\_ (thinking) — Ma...«cannonata» in senso bellico...O calcistico? —

Problema della polisemia

👺 Simone Scannapi	eco
-------------------	-----

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		

# WORD2VEC



- n
- Polisemia: confusione con parole utilizzabili con diversi significati in contesti diversi
- Contesti insufficienti: parole rare potrebbero non avere abbastanza esempi

>\_ (thinking) — Non ho molto capito il senso delle parole «supercazzola», «scappellamento» e «Antani»...—

Amici miei, 1975

:	Simone	Scanna	nieco
	OIITIOTIC	Ocumina	PICCO

Spring AI - Corso base

m Venis S.p.A, Venezia, IT

Note		