dgroove
Tecnologia emozionale.

# ICT Training Center

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



Note

# Language Artificial Intelligence

Note

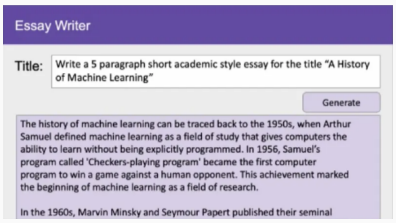
# Language AI
## ...o NLP

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



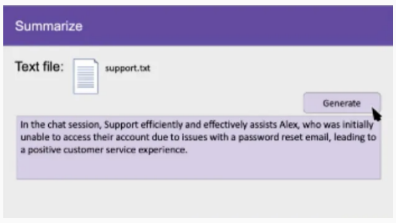
Note

Traduzione *context-sensitive*

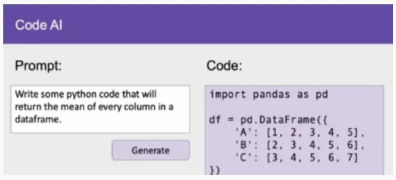



Generazione di testo

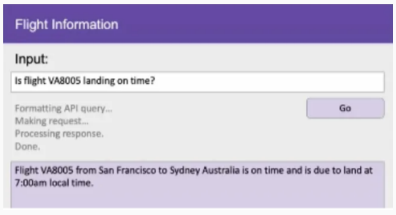
Estrazione di entità denominate




Riassunto di testi




Generazione di codice




Assistente digitale

Gara GPT vs Claude su traduzione *context-sensitive*



Note

# 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. . .
- . . . calcolando probabilità su ogni singola parola di un dizionario
- 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***



Note

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

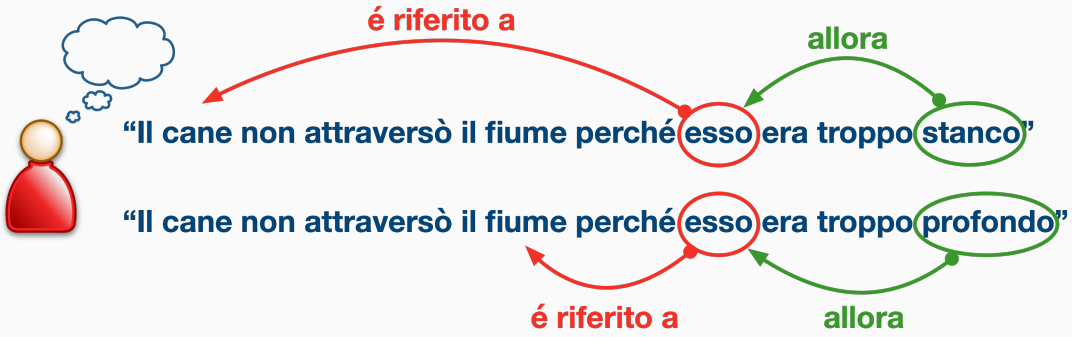
- Modelli che simulano la comprensione e la generazione del linguaggio umano con approcci statistici e modelli della rete neurale
- Predicono la parola successiva in una sequenza in base al contesto. . .
- . . . calcolando probabilità su ogni singola parola di un dizionario
- Qualunque *task* NLP può 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***



Note




- Il 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?!



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



Note

- 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**



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 ___?''
...
```



Note

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

### Esempi di *pattern matching* ELIZA

```
''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**



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**



Note

# Bag-of-Words
## Rappresentazione testo non strutturato

- Menzionato per la prima volta negli anni ’50, popolare negli anni 2000
- Metodo per rappresentare il testo non strutturato in formato numerico
- Il linguaggio è complicato per i calcolatori
  - Il testo perde significato quando rappresentato da 0 e 1
- *Focus* principale: rappresentare il linguaggio in modo strutturato per l’uso da parte dei calcolatori



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*
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

| Token | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| che | bel | cane | che | hai | ma | il | mio | gatto | è | più | bello |

1 *"Che bel cane che hai!"*
2 *"Ma il mio gatto è più bello"*

**Tokenizzazione**



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*
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

**Vocabolario**

| che | bell- | can- | avere | ma | il | mi- | gatt- | essere | più | bell- |
|---|---|---|---|---|---|---|---|---|---|---|

1 *"Che bel cane che hai!"*
2 *"Ma il mio gatto è più bello"*

**Creazione del vocabolario**



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*
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

**Vocabolario**

| | che | bell- | can- | avere | ma | il | mi- | gatt- | essere | più | bell- |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

1 *"Che ␣ bel ␣ cane ␣ _____ ␣ che ␣ hai!"*
2 *"Ma ␣ il ␣ _____ ␣ mio ␣ gatto ␣ è ␣ più ␣ _____ ␣ bello"*

**Vettori *bag-of-words***



Note

- 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

| | Vocabolario | | | |
|---|---|---|---|---|
| | Simone | mangiare | l- | insalat- |
| $1$ | 1 | 1 | 1 | 1 |
| $2$ | 1 | 1 | 1 | 1 |

$1$ *"Simone mangiò l' insalata"*
$2$ *"L' insalata mangiò Simone"*

**Problema dell'ordine delle parole**



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

> *"Concetti come «Re» e «Regina» dovrebbero essere correlati…"*
> *"Sinonimi come «felice» e «gioioso» dovrebbero avere una rappresentazione simile…"*
> *"E con parole come «pitone» e «serpente», una generalizzazione dell'altra?!"*
>
> **Mancanza di relazioni semantiche**



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 $10^5$?!"*
>
> **Problema della dimensionalità**



Note

# Word Embeddings (2013)
## Word2Vec e le rappresentazioni dense

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



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 [...]"*
2 *"[...] Il veterinario ha deciso di sterilizzare il gatto [...]"*
3 *"[...] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario [...]"*
4 *"[...] 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**



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 […]"*
2. *"[…] Il veterinario ha preferito sterilizzare il gatto […]"*
3. *"[…] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario […]"*
4. *"[…] 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**



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



Note

- 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**



Note

- 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**



Note