

Corso di laurea magistrale in Data Science
Anno Accademico 2021-2022

Relatore: Matteo Palmonari

Correlatori: Riccardo Pozzi & Manuel Vimercati

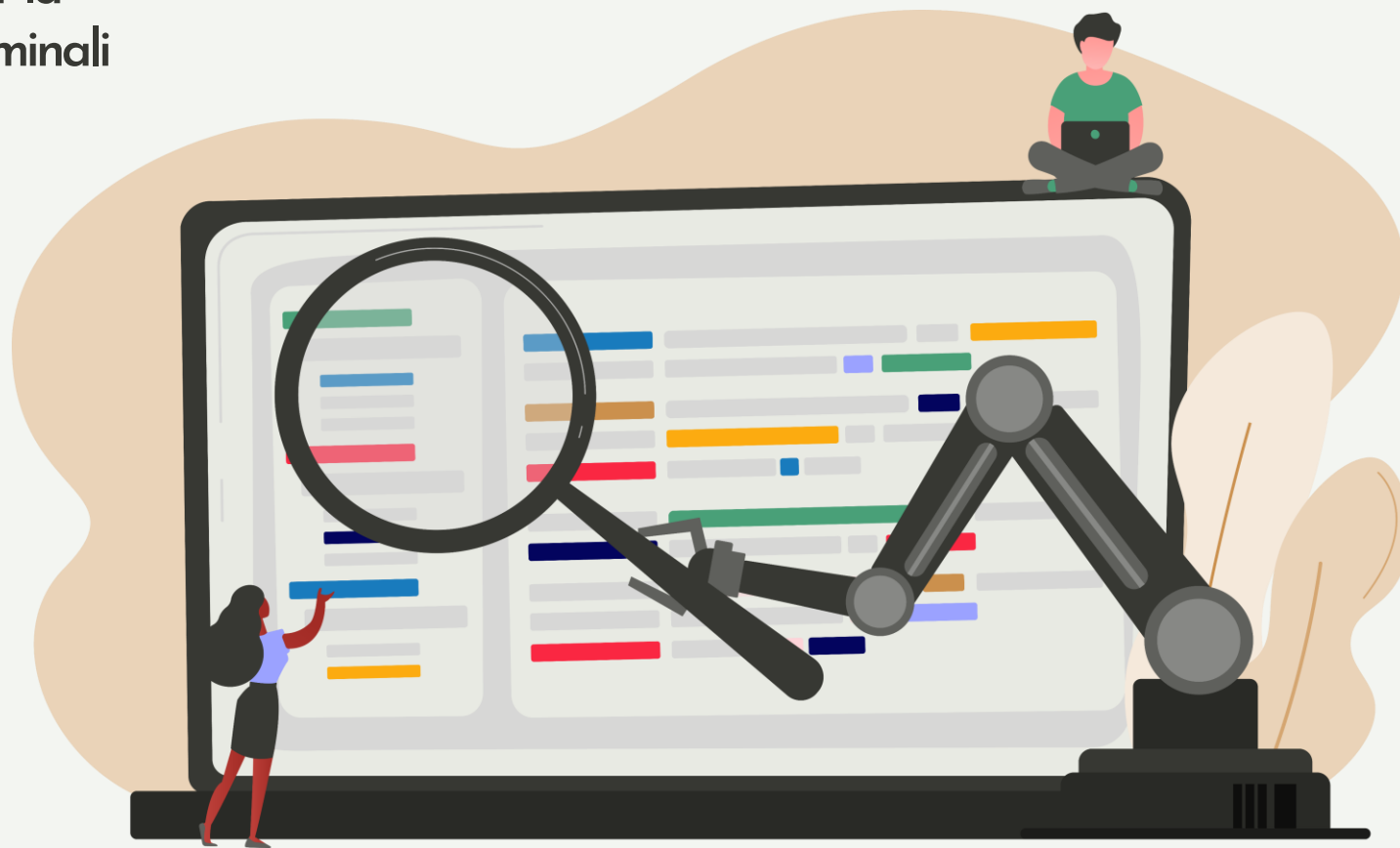
PROMET

Un approccio adattivo prompt-based per la
classificazione fine di menzioni di entità nominali

Tesi a cura di Riccardo Rubini



UNIVERSITÀ DEGLI STUDI MILANO-BICOCCA
Dipartimento di Informatica, Sistemistica e Comunicazione



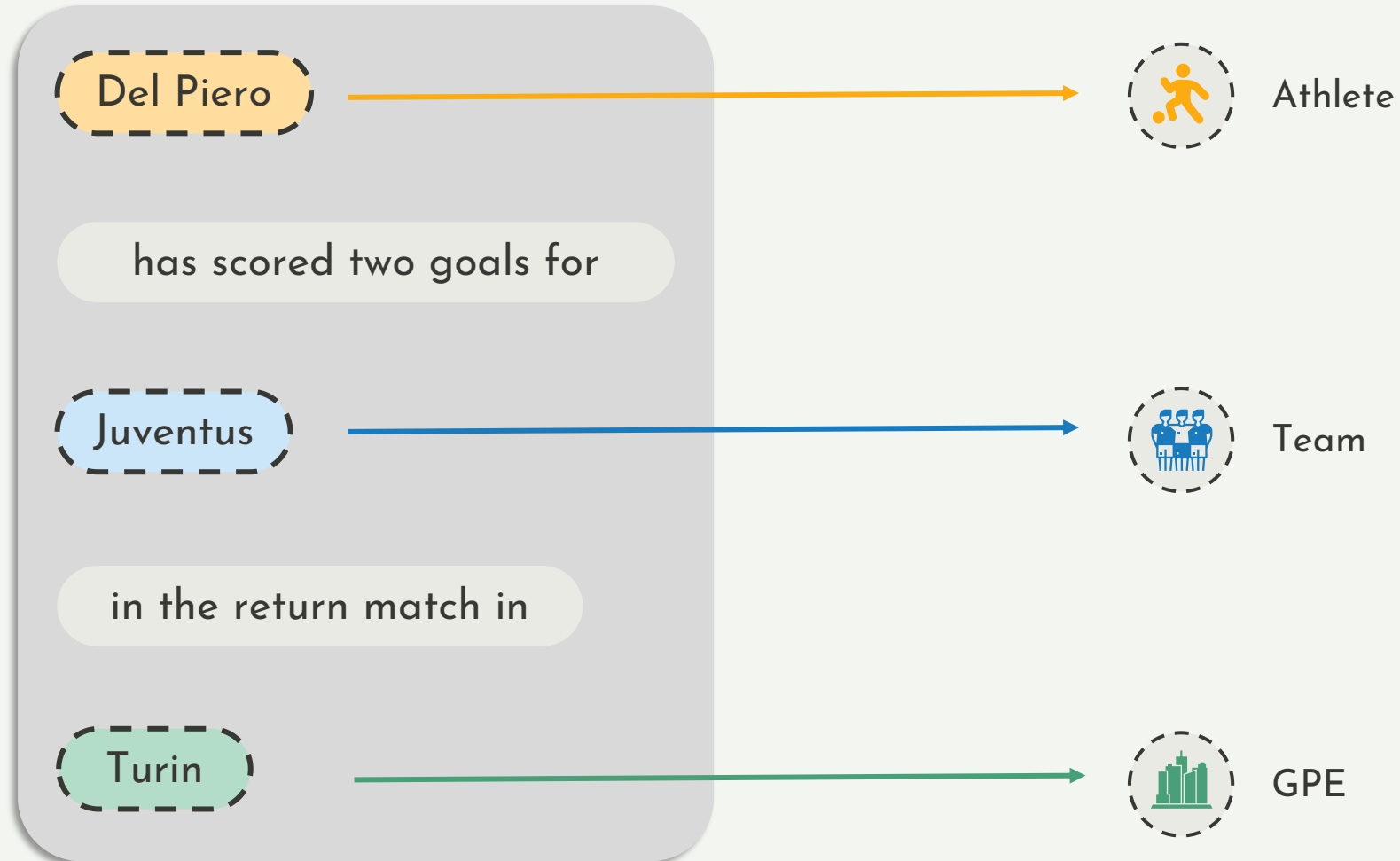
Classificazione fine delle entità nominali

Introduzione

Del Piero has scored two goals for Juventus in the return match in Turin.

Classificazione fine delle entità nominali

Introduzione

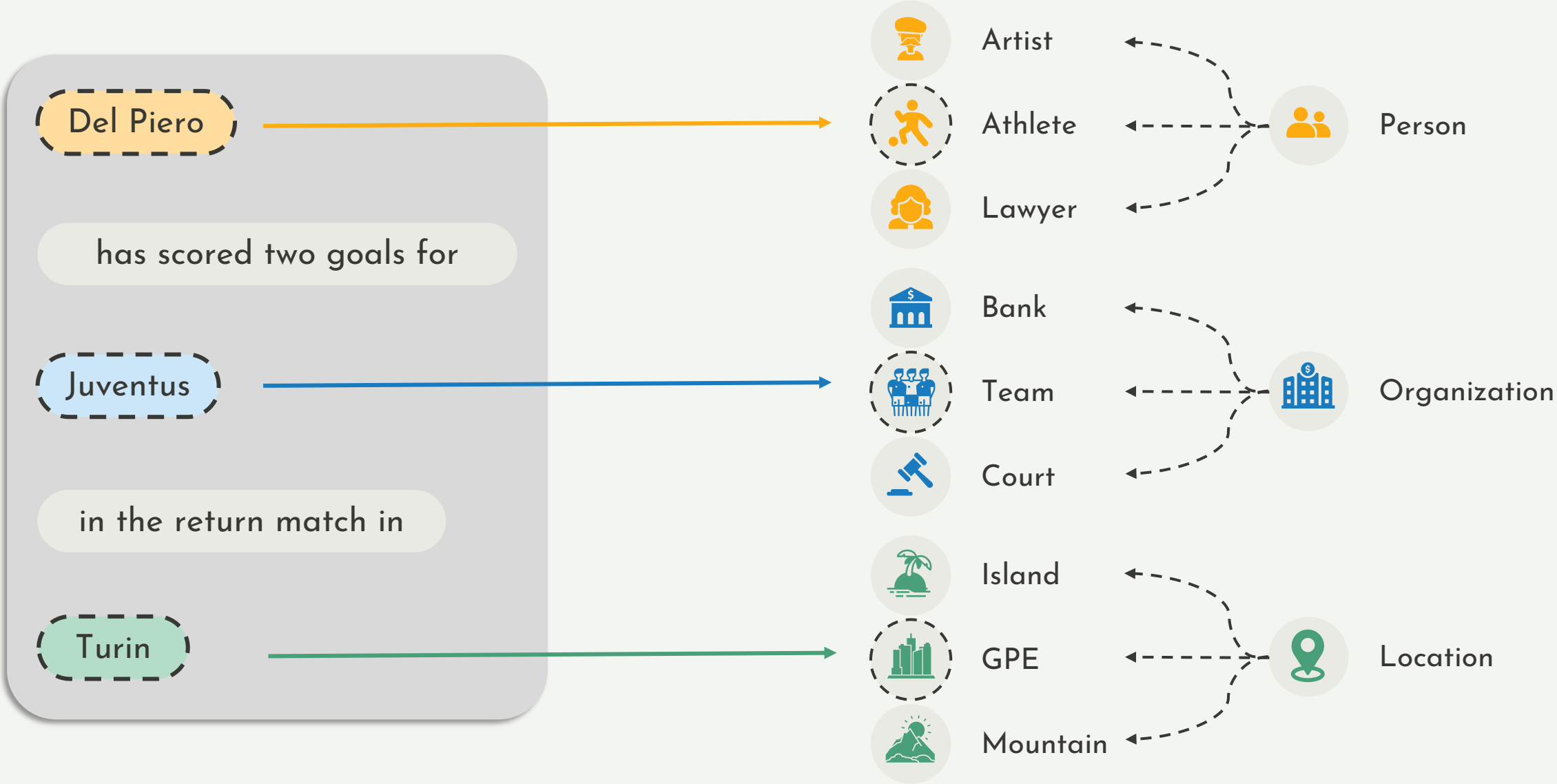


La classificazione fine dell'entità (in inglese **Fine-Grained Entity Typing, FET**) consiste nell'**assegnare** a ciascuna entità nominale presente nel testo **un'etichetta specifica** tra un set predefinito.

Classificazione fine delle entità nominali

Introduzione

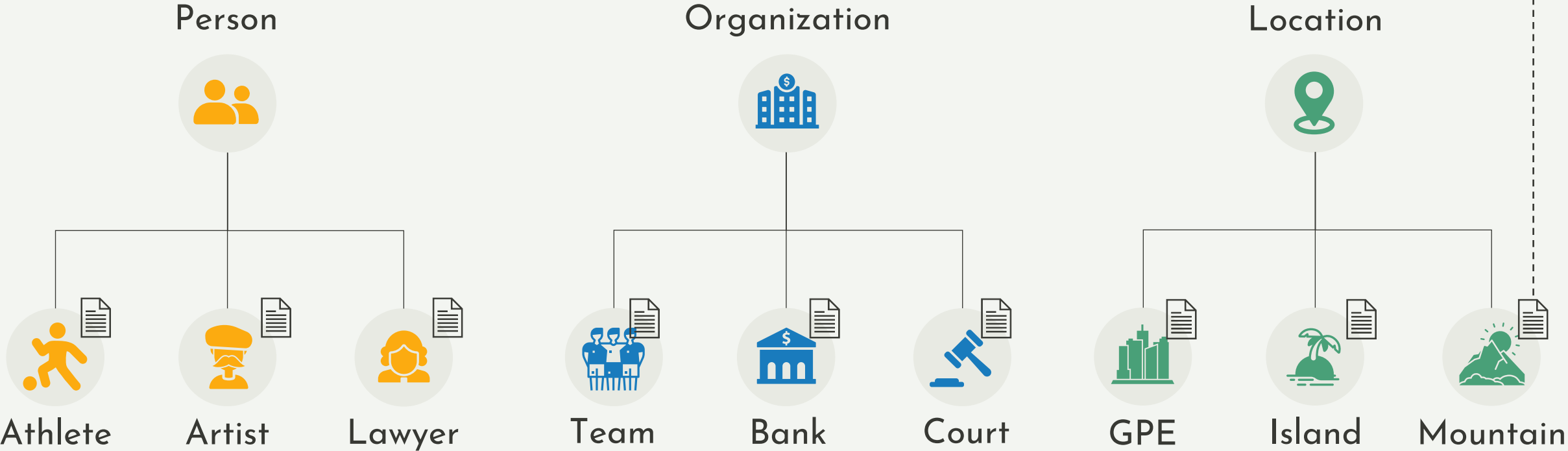
Nel FET le categorie sono **organizzate in un'ontologia**, all'interno della quale sono definite delle **relazioni gerarchiche**.



Classificazione fine delle entità nominali

Caso d'uso classico VS Specializzazione dell'Ontologia

Fin dall'inizio si hanno a disposizione osservazioni per ogni etichetta specifica presente nel set di classificazione ←



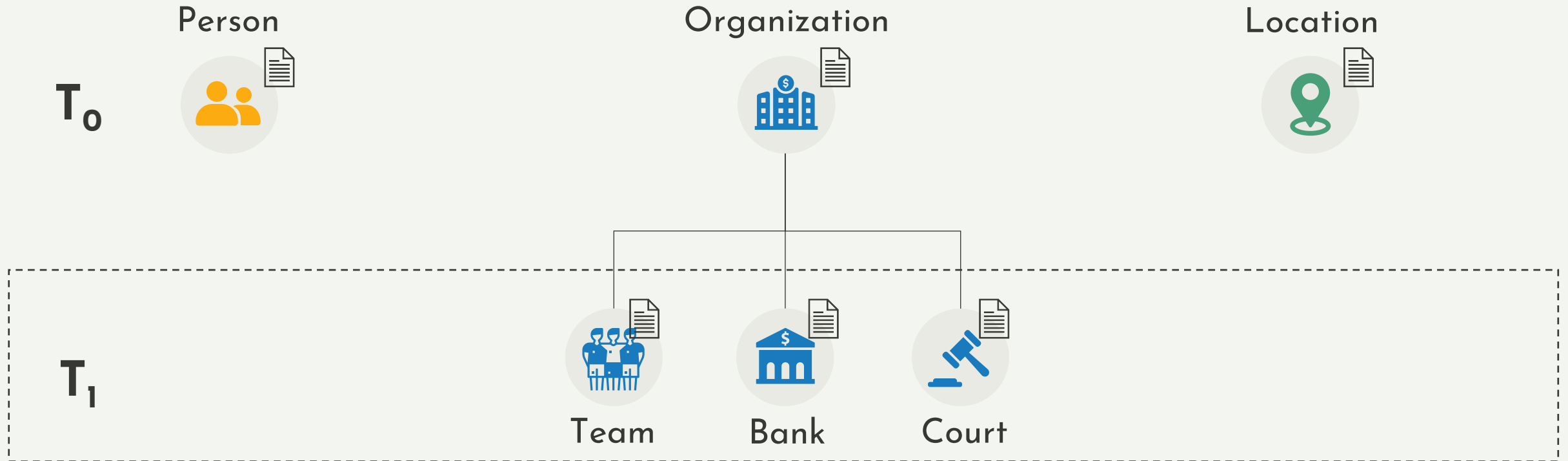
Classificazione fine delle entità nominali

Caso d'uso classico VS **Specializzazione dell'Ontologia**



Classificazione fine delle entità nominali

Caso d'uso classico VS **Specializzazione dell'Ontologia**



Al tempo T_1 vengono aggiunte osservazioni annotate con le etichette specifiche dei tipi padre che si vogliono specializzare (ovvero aggiungere tipi figlio)

Classificazione fine delle entità nominali

Adattività dei modelli

**Tramite la specializzazione
dell'ontologia il modello può adattarsi
a scenari applicativi diversi**



Infatti l'etichettatura dei tipi padre è condivisa tra i vari casi d'uso data la sua generalità. Ogni scenario applicativo però necessita di classi specifiche, le quali differiscono da caso d'uso a caso d'uso.



Classificazione fine delle entità nominali

Scenario Few-Shot



Classificazione fine delle entità nominali

Scenario Few-Shot



Bassa frequenza delle
entità specifiche nel testo



Classificazione fine delle entità nominali

Scenario Few-Shot



Bassa frequenza delle entità specifiche nel testo



Ambiguità tra le entità di tipi semanticamente simili



Classificazione fine delle entità nominali

Scenario Few-Shot



Bassa frequenza delle entità specifiche nel testo



Ambiguità tra le entità di tipi semanticamente simili



I costi per annotazione sono elevati



Obiettivi di PROMET

1

Modello per la classificazione fine di entità con etichette organizzate in una gerarchia

2

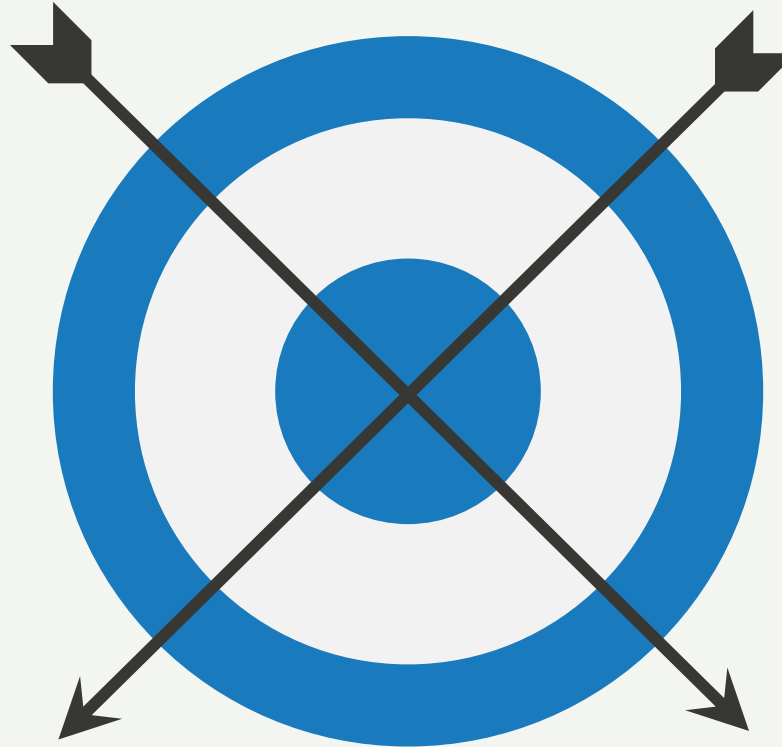
Modello in grado di lavorare bene anche in situazioni di scarsità di dati

3

Modello semplice e in grado di adattarsi ai vari casi d'uso tramite la specializzazione dell'ontologia

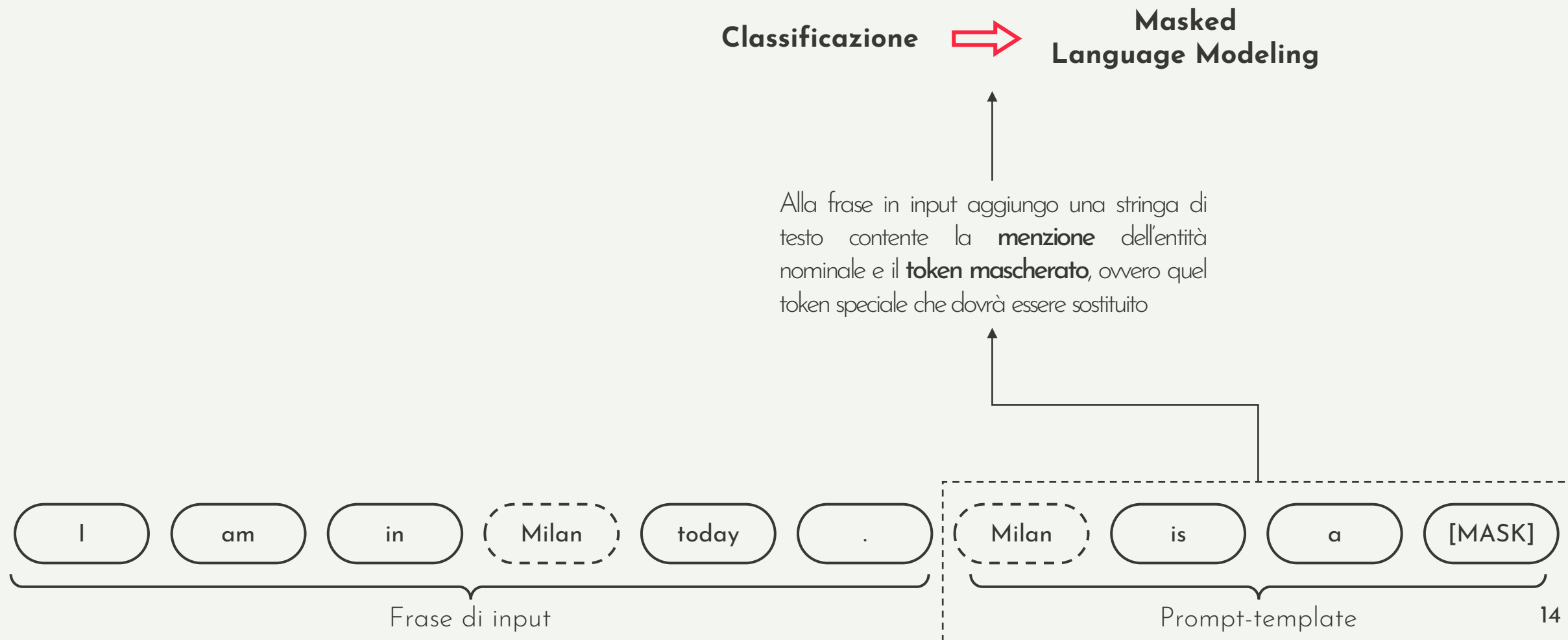
4

Valutare il modello su dati benchmark per confrontarlo con lo stato dell'arte



Stato Dell'Arte

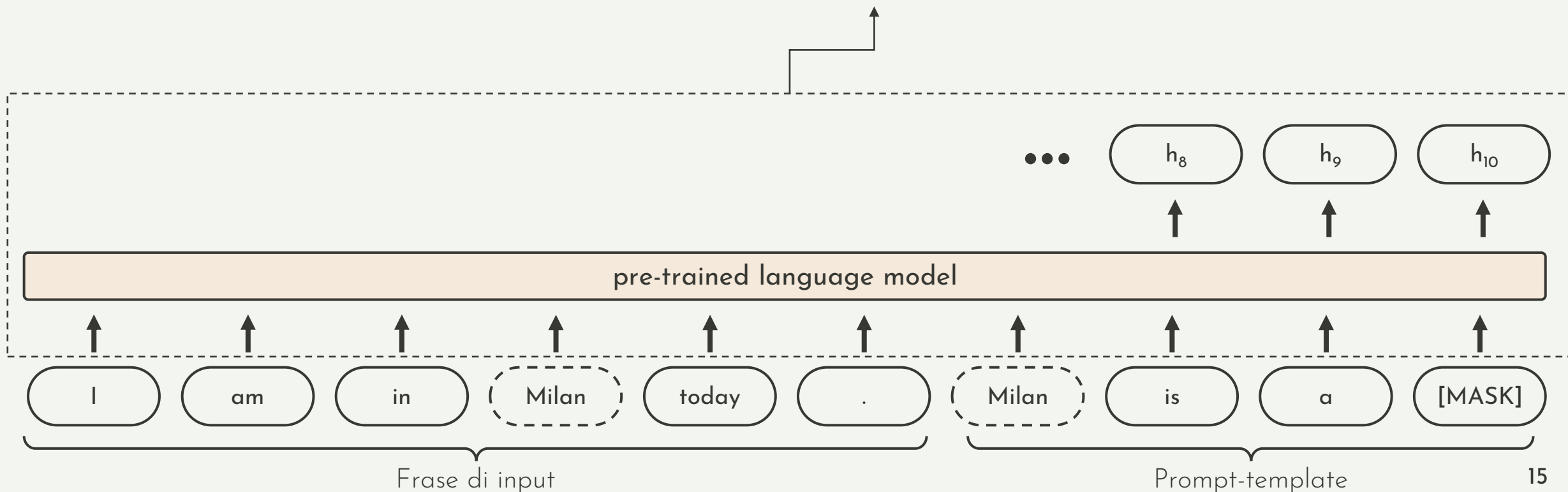
Prompt Learning



Stato Dell'Arte

Prompt Learning

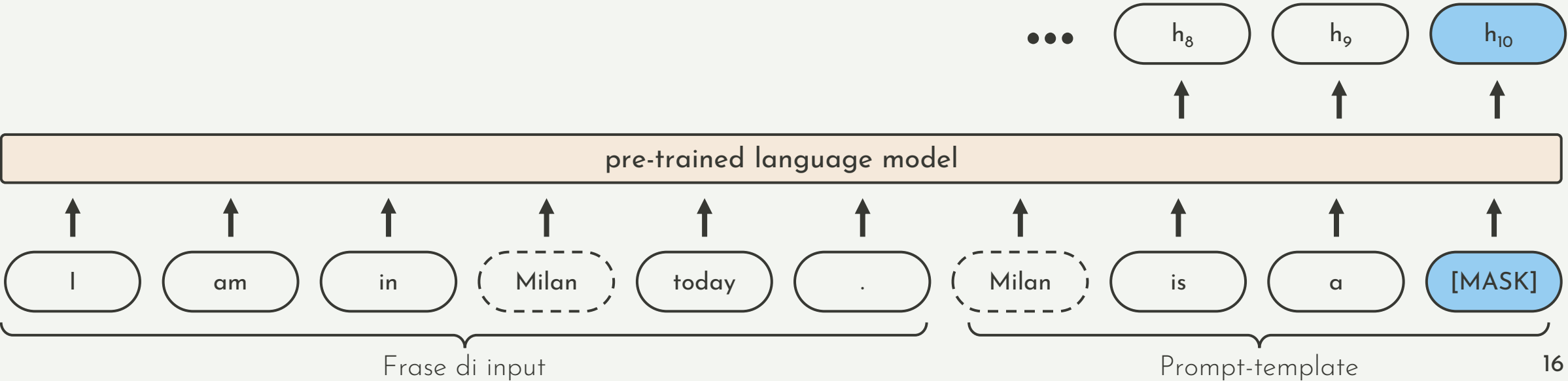
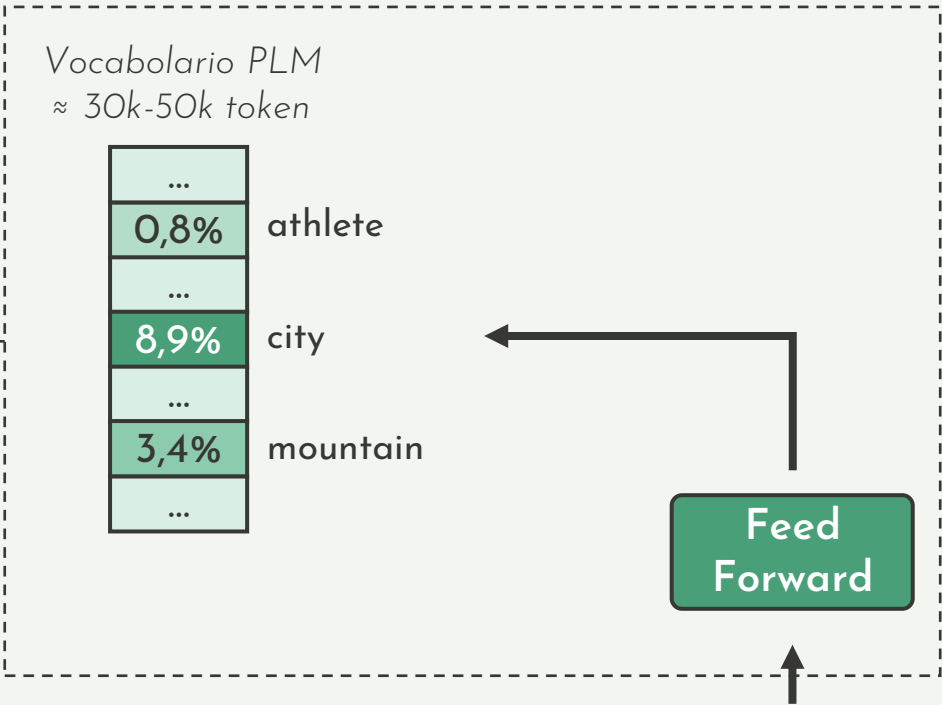
La **sequenza di token** viene **processata** da un **pre-trained language model (PLM)**, ovvero un **modello** linguistico **allenato su milioni di dati**. Esempi di PLM sono BERT e RoBERTa. Questi sfruttando tecniche di **apprendimento profondo** e meccanismi di **auto-attenzione** produce **word embedding (h)** per ogni token della sequenza in input.



Stato Dell'Arte

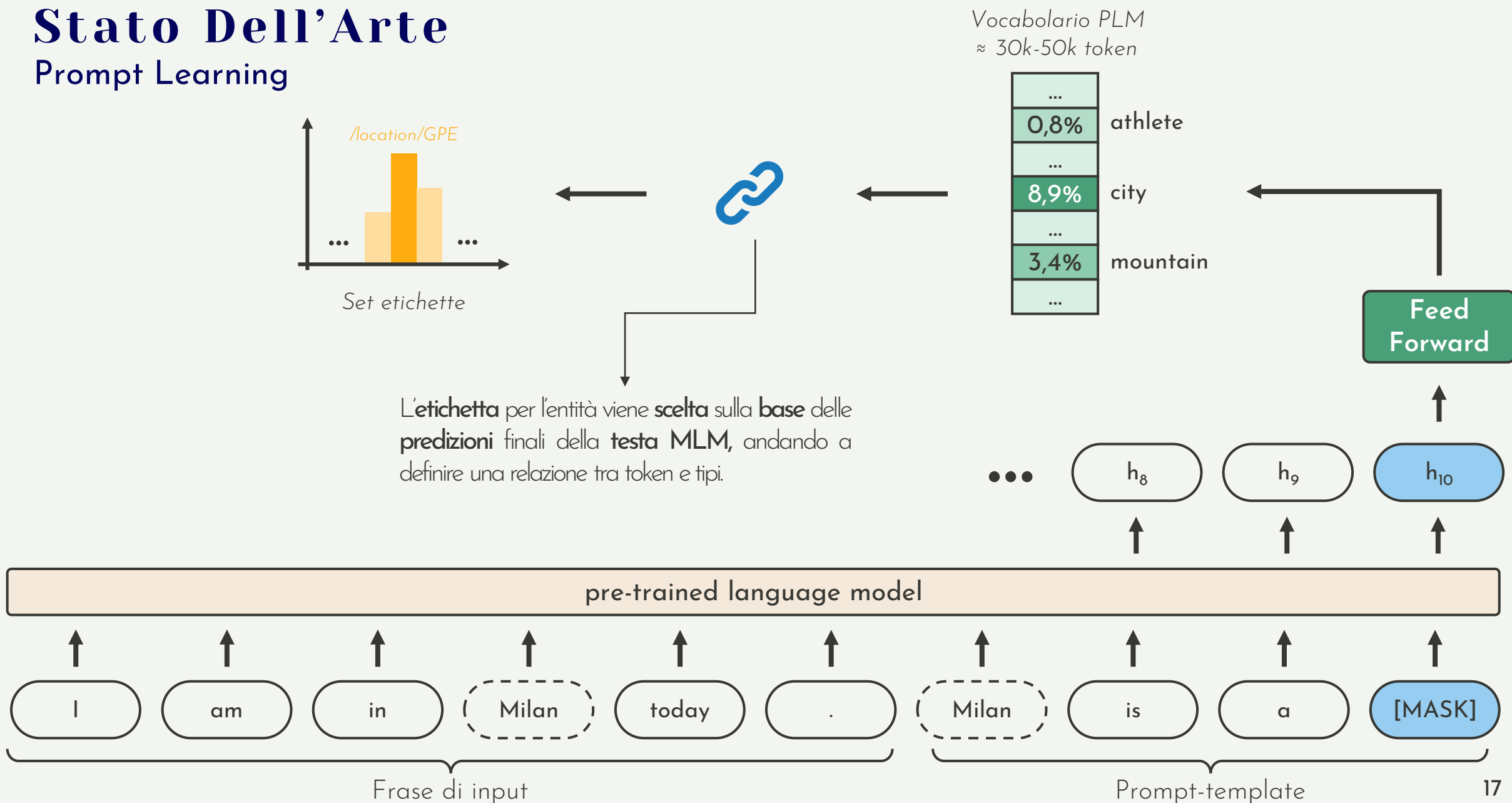
Prompt Learning

La **rete feed-forward** serve per eseguire la task di masked language modeling (MLM). Questa produce la **distribuzione di probabilità** sull'intero vocabolario, **indicando i token più verosimili in sostituzione** di quello mascherato.



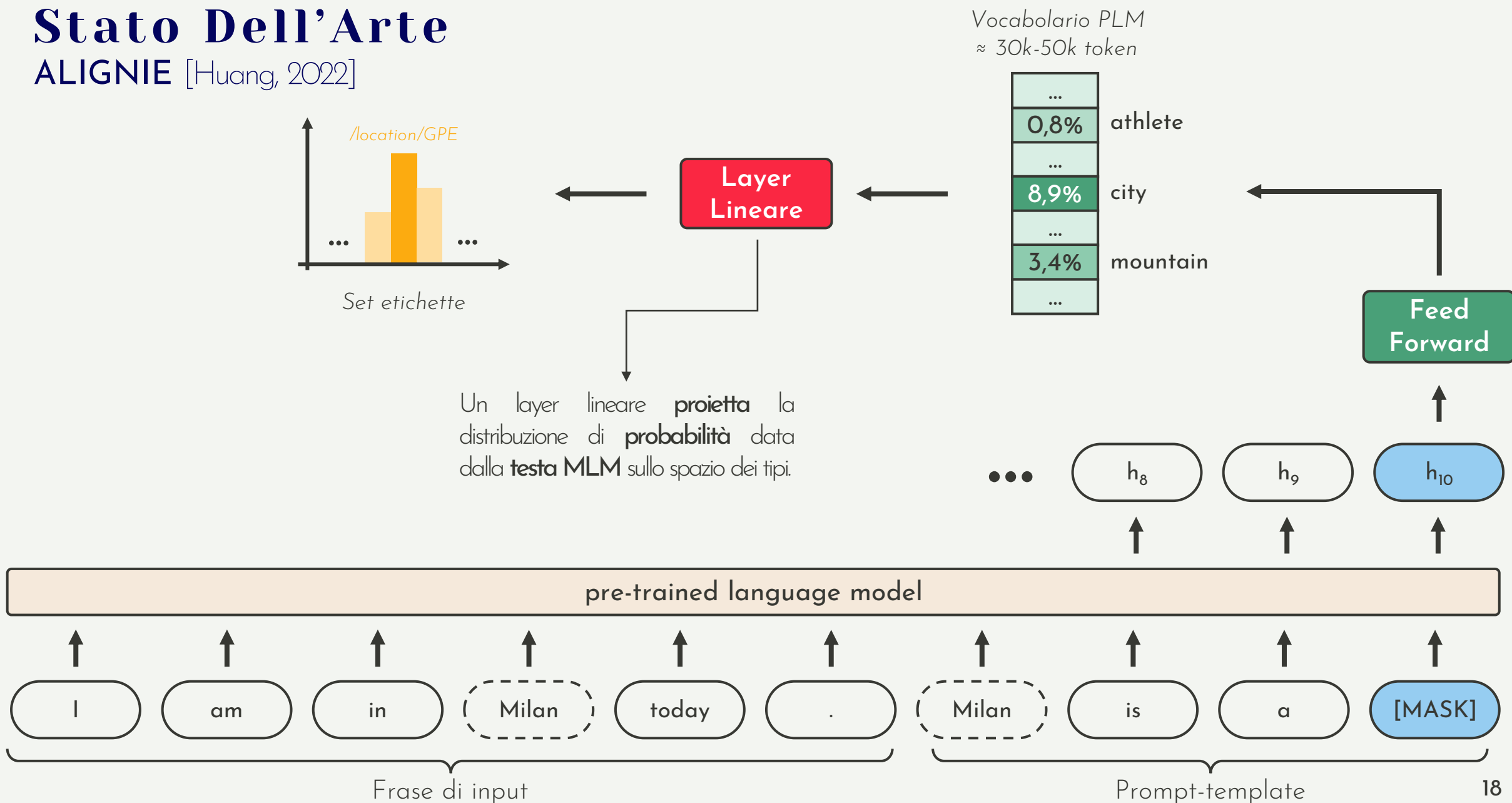
Stato Dell'Arte

Prompt Learning



Stato Dell'Arte

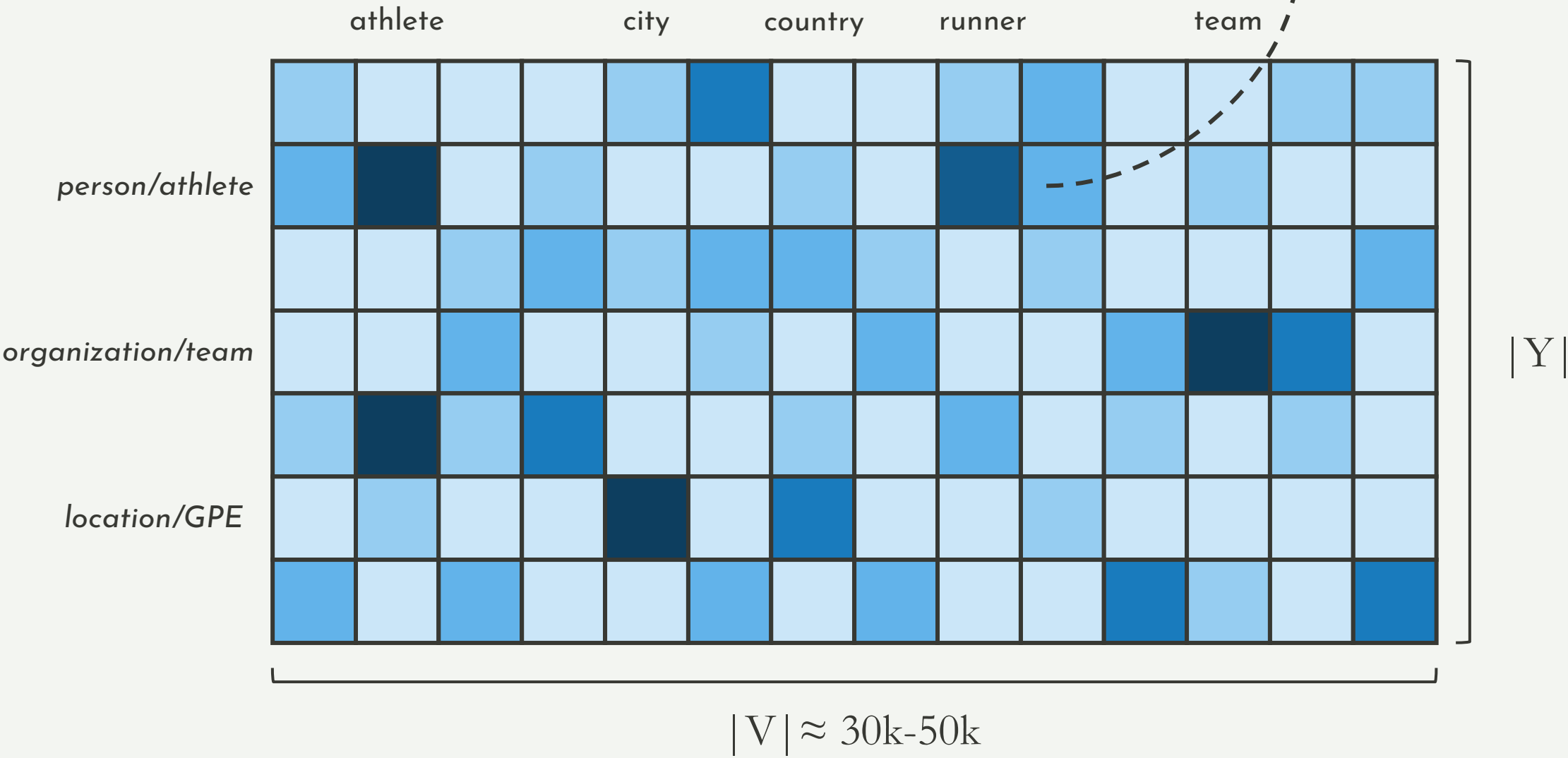
ALIGNIE [Huang, 2022]



Stato Dell'Arte

Addestramento pesi di ALIGNIE [Huang, 2022]

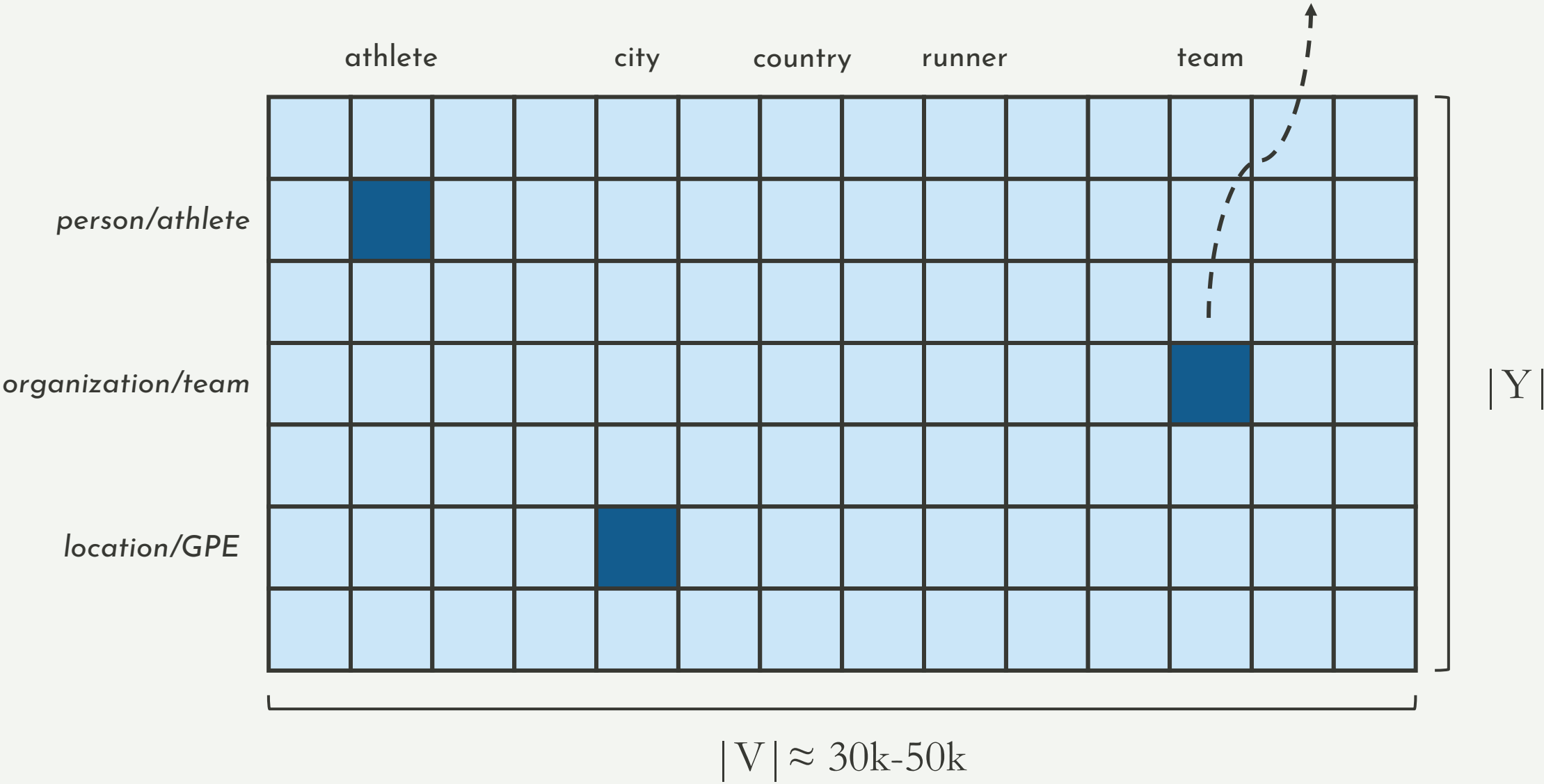
I pesi del layer di proiezione tra i token e i tipi vengono **aggiornati ed ottimizzati** durante la fase di **training**



Stato Dell'Arte

Inizializzazione dei pesi di ALIGNIE [Huang, 2022]

I pesi del layer lineare vengono
inizializzati manualmente per evitare
l'overfitting, dando peso maggiore ad
almeno una parola chiave per tipo



Stato Dell'Arte

Selezione parole chiave in ALIGNIE [Huang, 2022]

**Selezionare manualmente le parole chiave
è un compito impegnativo e oneroso...**



Stato Dell'Arte

Selezione parole chiave in ALIGNIE [Huang, 2022]

**Selezionare manualmente le parole chiave
è un compito impegnativo e oneroso...**



- 1** Alcuni tipi possono non essere ben definiti dal punto di vista semantico



Stato Dell'Arte

Selezione parole chiave in ALIGNIE [Huang, 2022]

**Selezionare manualmente le parole chiave
è un compito impegnativo e oneroso...**



- 1** Alcuni tipi possono non essere ben definiti dal punto di vista semantico
- 2** Alcune etichette posso essere disambiguate solo da parole chiave composte da più token



Stato Dell'Arte

Selezione parole chiave in ALIGNIE [Huang, 2022]

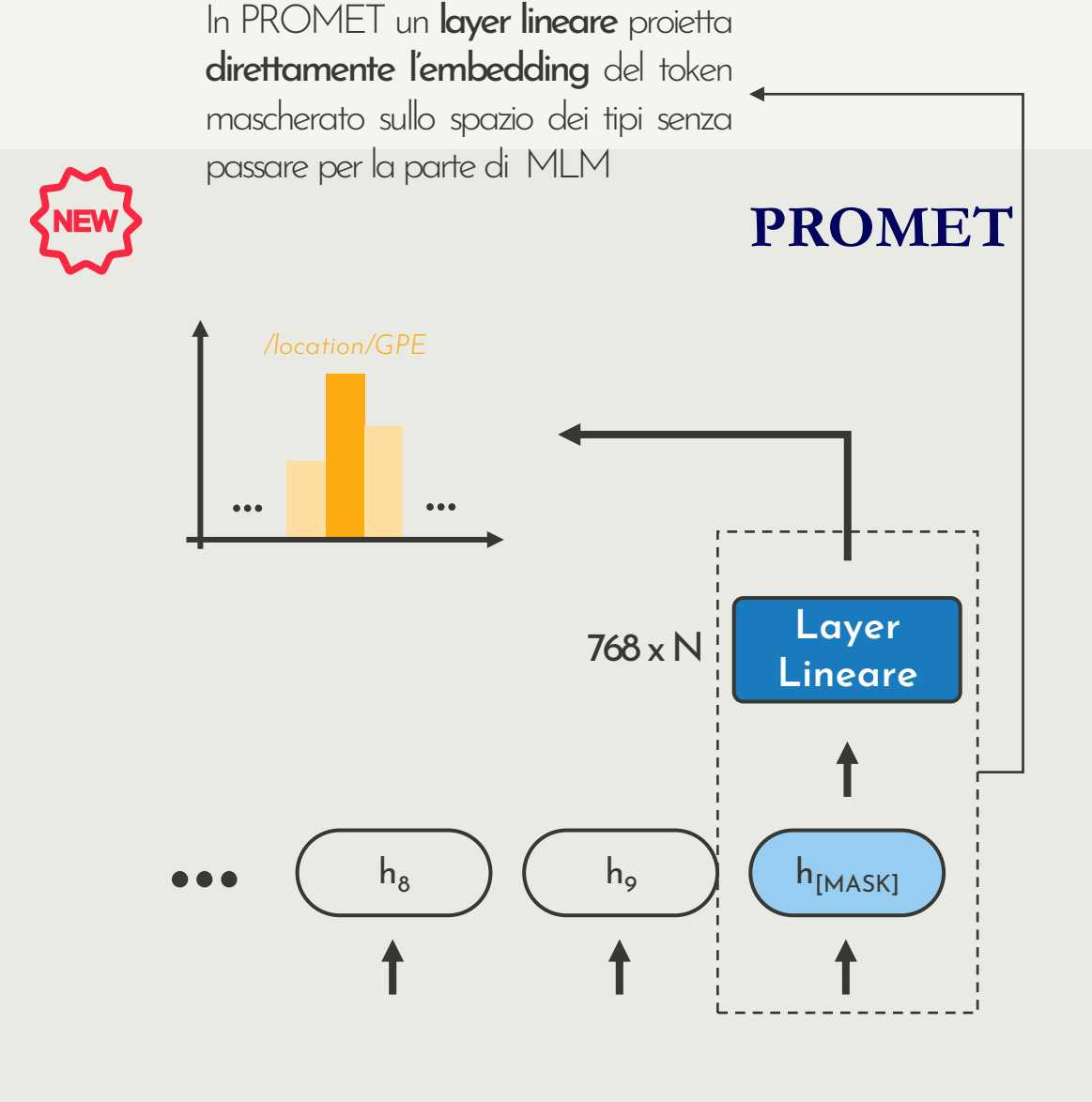
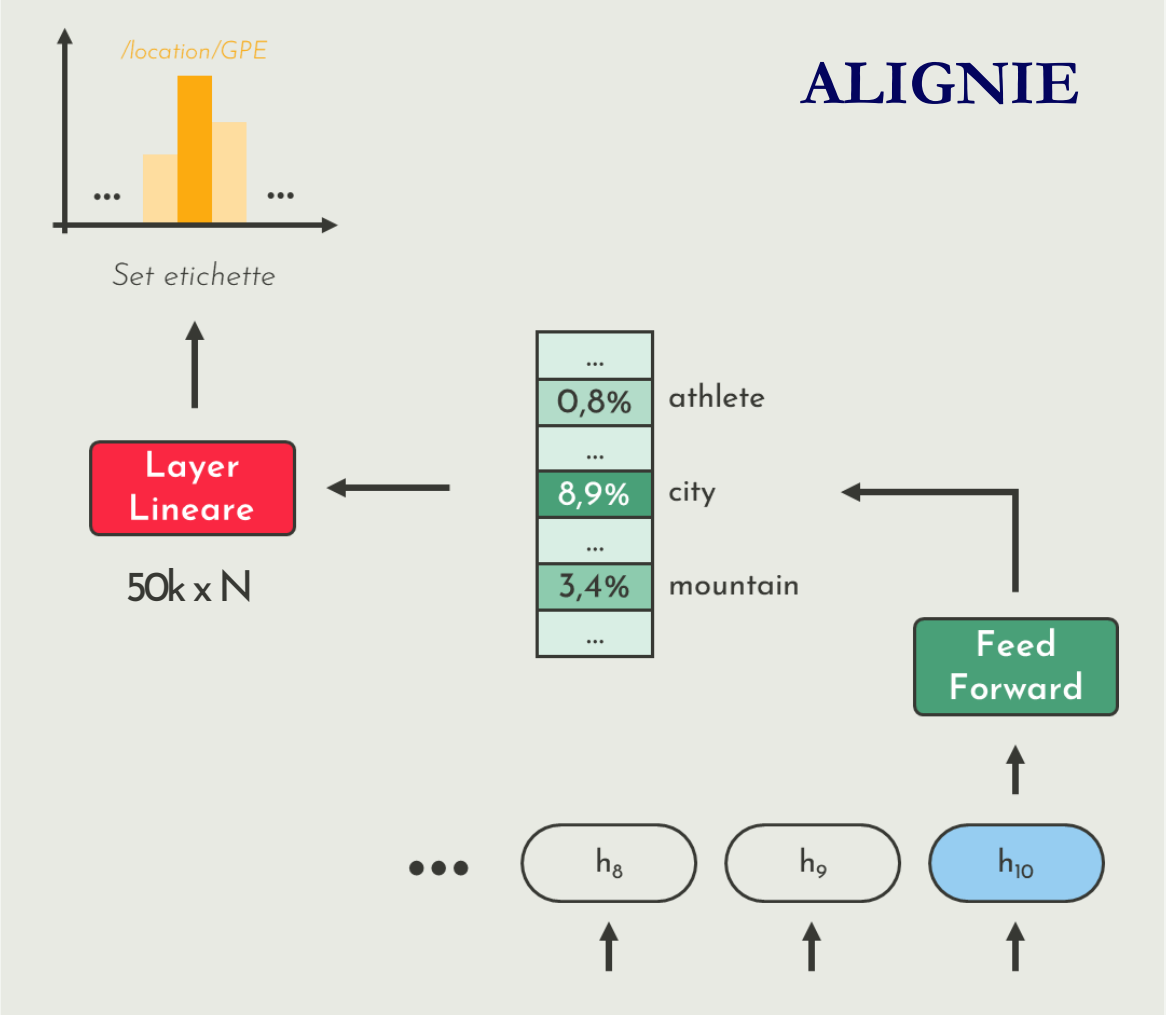
**Selezionare manualmente le parole chiave
è un compito impegnativo e oneroso...**



- 1** Alcuni tipi possono non essere ben definiti dal punto di vista semantico
- 2** Alcune etichette posso essere disambiguate solo da parole chiave composte da più token
- 3** Alcuni casi d'uso possono richiedere centinaia o migliaia di etichette



PROMET vs ALIGNIE



PROMET vs ALIGNIE

**PROMET è più semplice degli altri approcci
prompt-based**



PROMET vs ALIGNIE

PROMET è più semplice degli altri approcci prompt-based



1

Il layer lineare ha un **numero di parametri di due ordini di grandezza inferiore** rispetto a quello di **ALIGNIE**



PROMET vs ALIGNIE

PROMET è più semplice degli altri approcci prompt-based



1

Il layer lineare ha un **numero di parametri di due ordini di grandezza inferiore** rispetto a quello di **ALIGNIE**

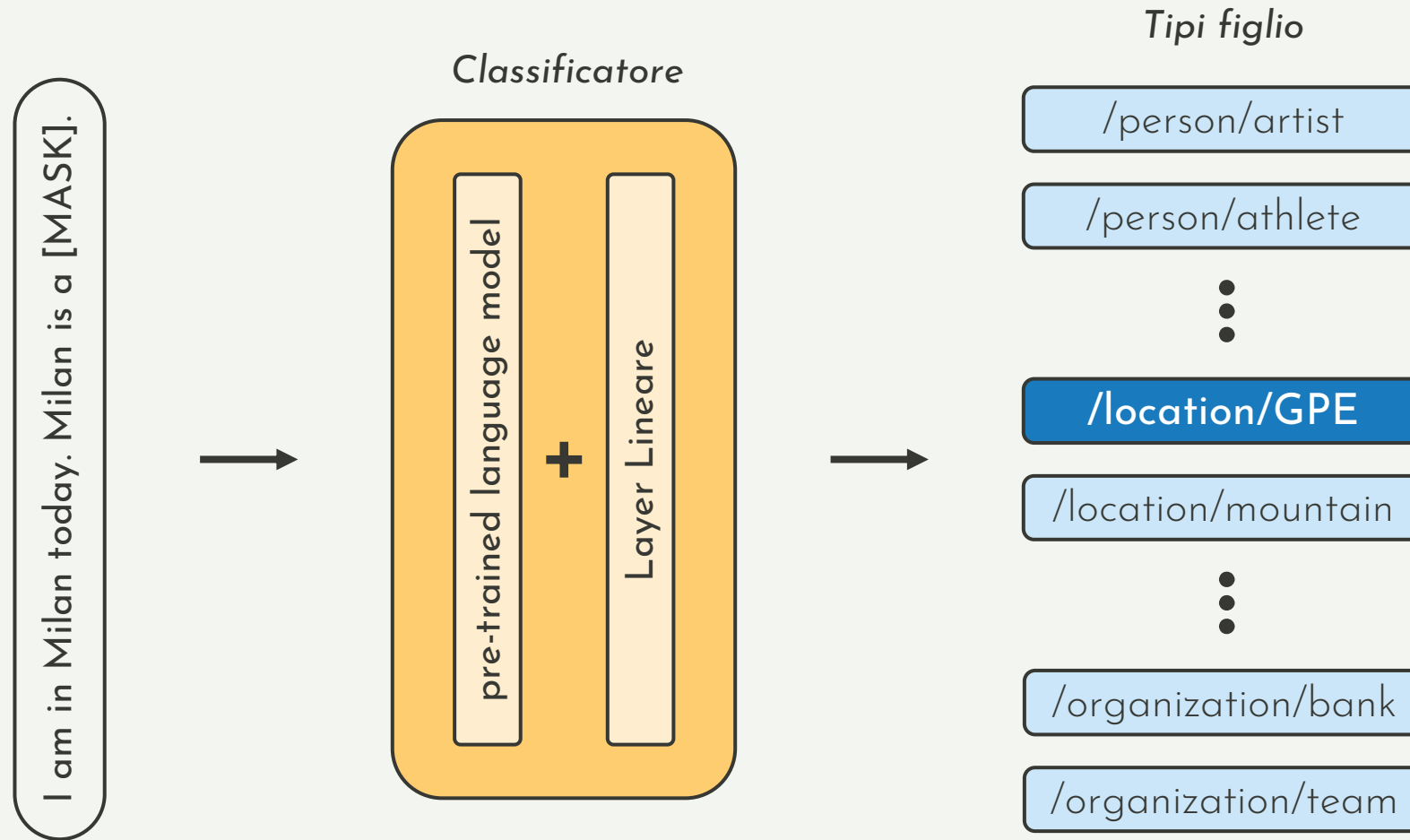
2

Non c'è bisogno di selezionare alcuna parola chiave, non essendo coinvolta la parte di MLM ed avendo un **rischio ridotto di overfitting**



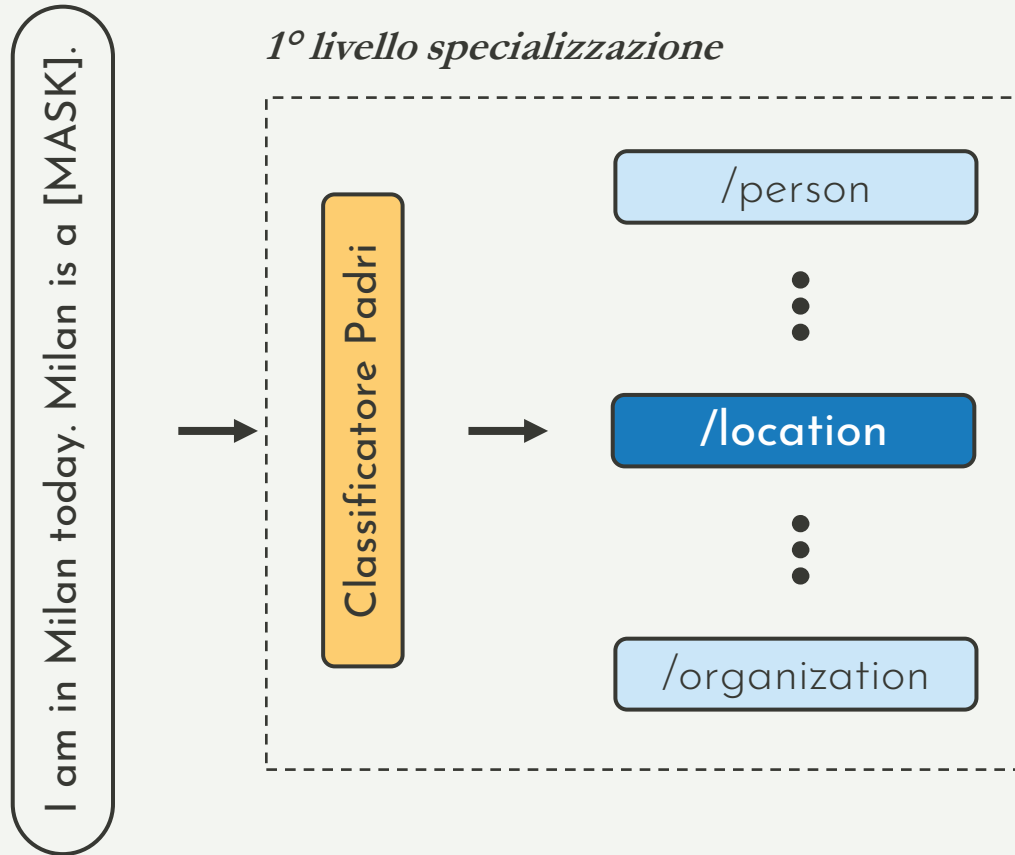
PROMET(F)

Modalità Flat di PROMET



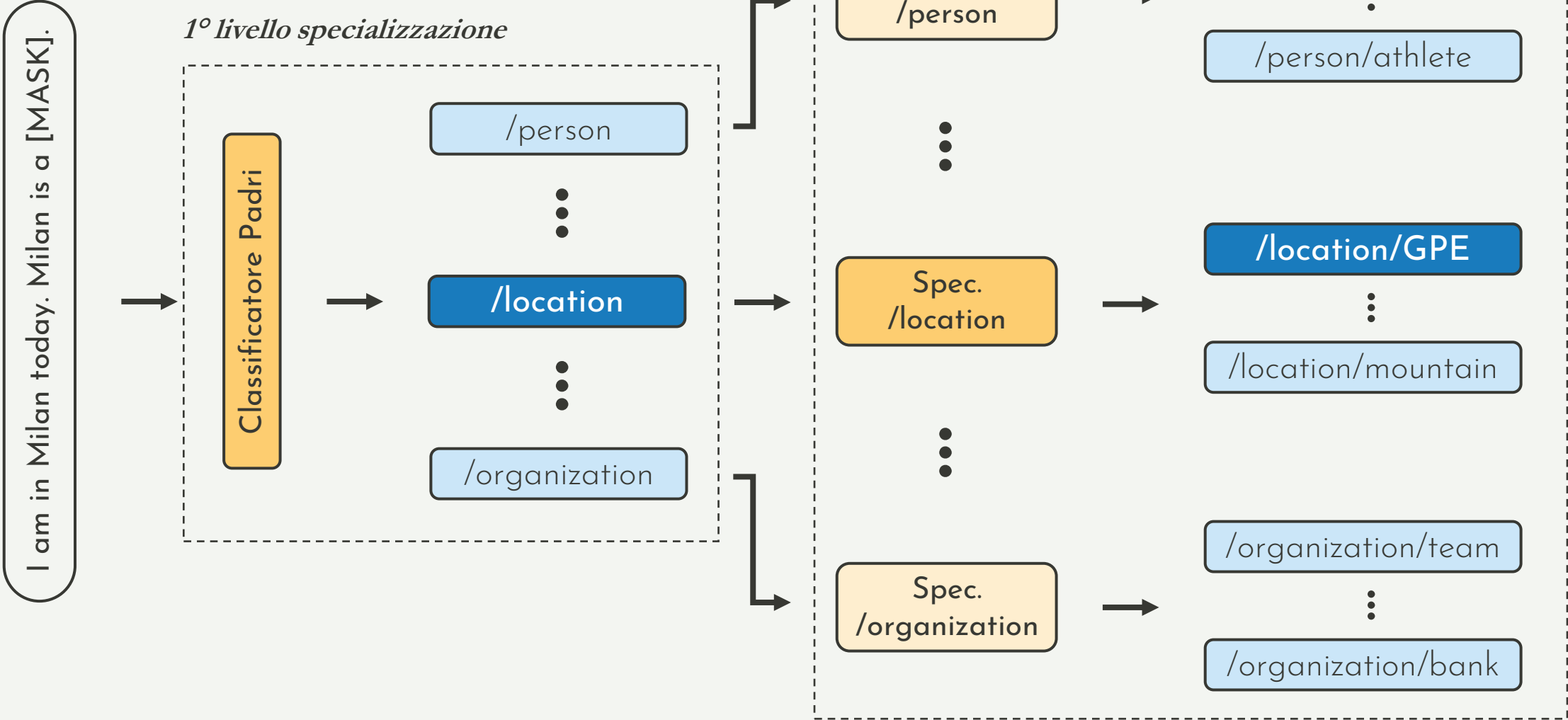
PROMET(S)

Modalità Stack di PROMET



PROMET(S)

Modalità Stack di PROMET

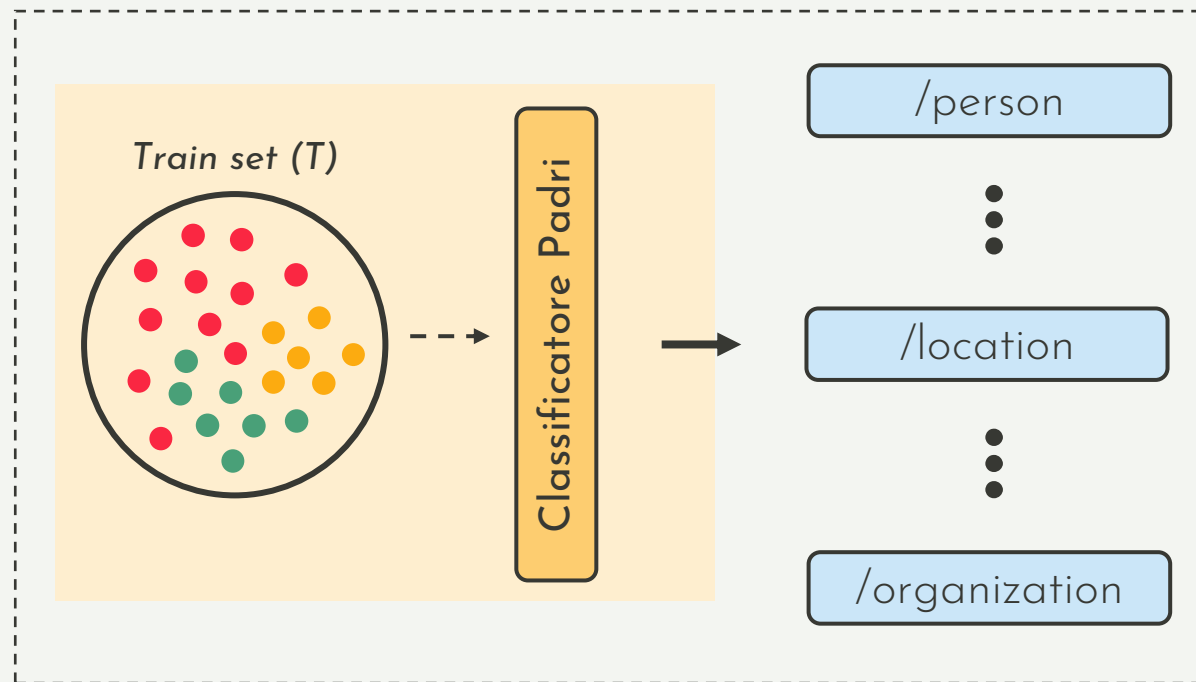


PROMET(S)

Addestramento della modalità Stack

1° livello specializzazione

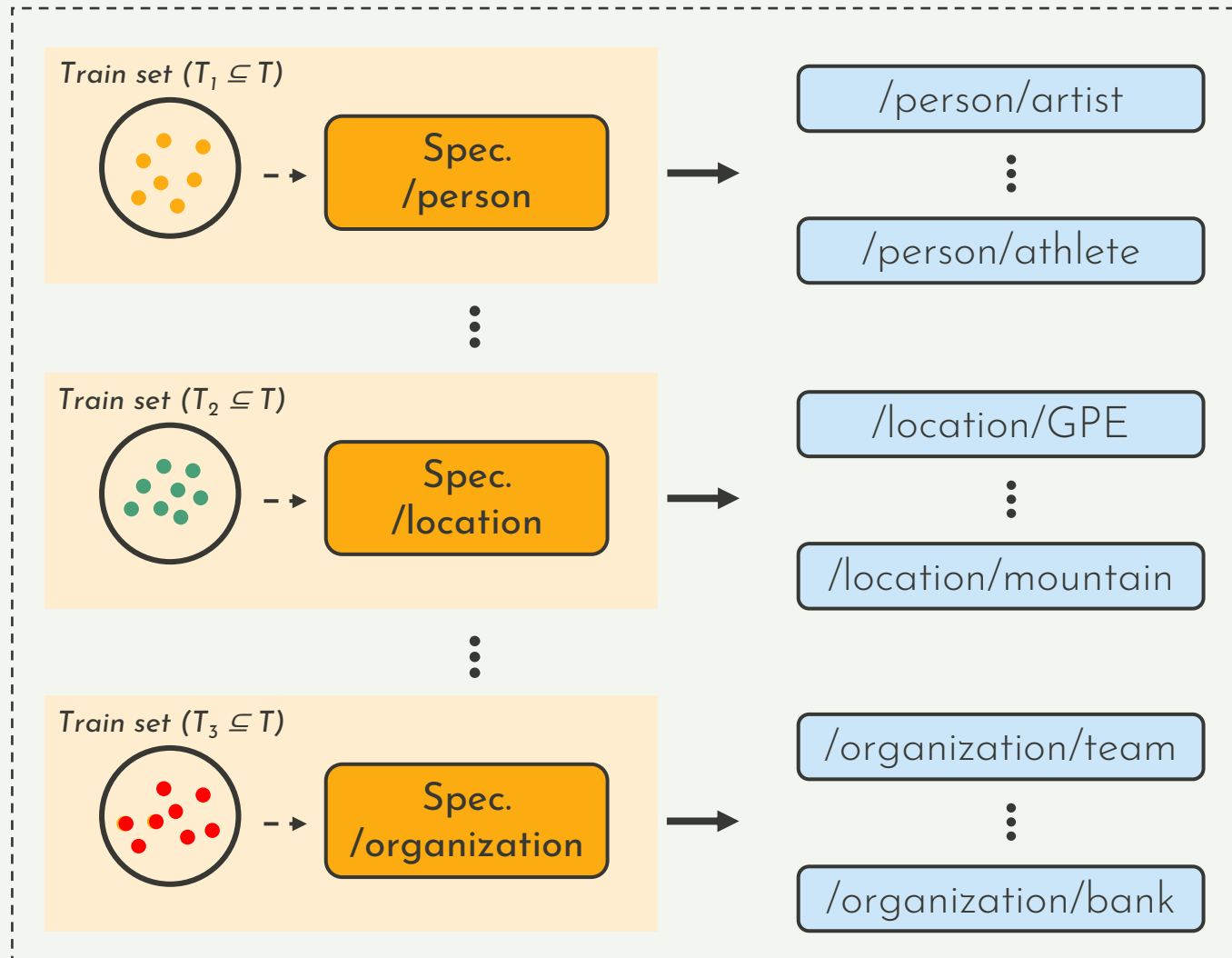
Il classificatore del primo livello è **allenato su tutti i dati** a disposizione



PROMET(S)

Addestramento della modalità Stack

2° livello specializzazione



Ogni classificatore del secondo livello è **allenato su un sottoinsieme dei dati** di training, contenete le osservazioni relative all'etichetta padre da specializzare

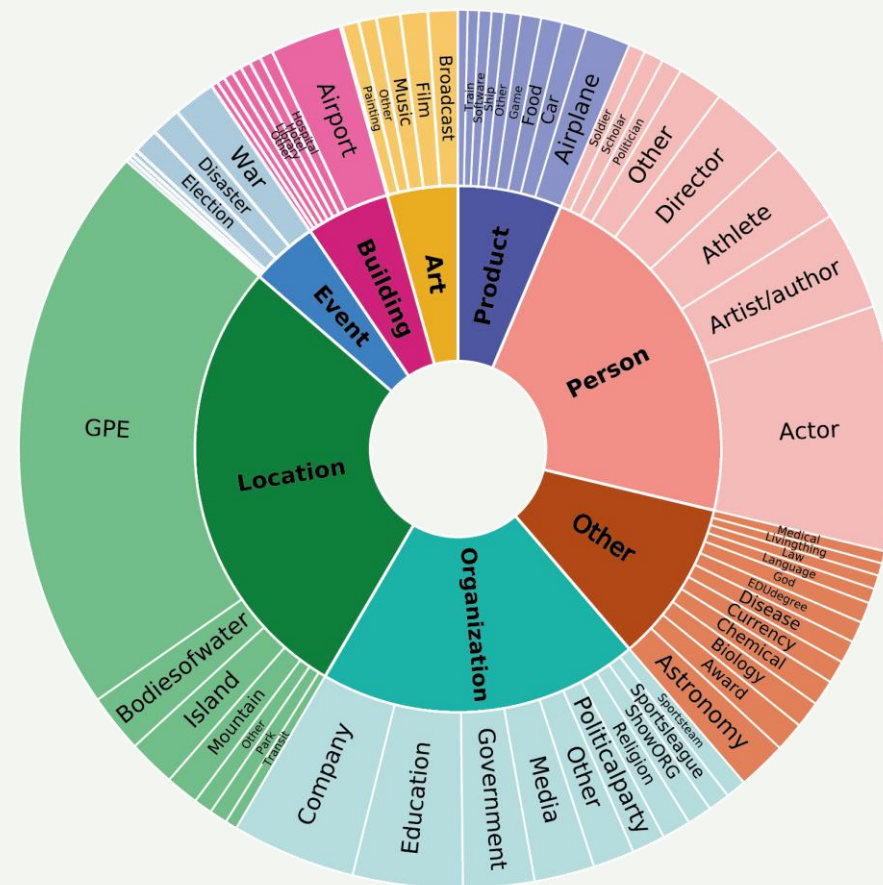
Esperimenti



Esperimenti

Dataset benchmark

Few-Nerd



Gerarchia e distribuzione dei tipi in Few-Nerd

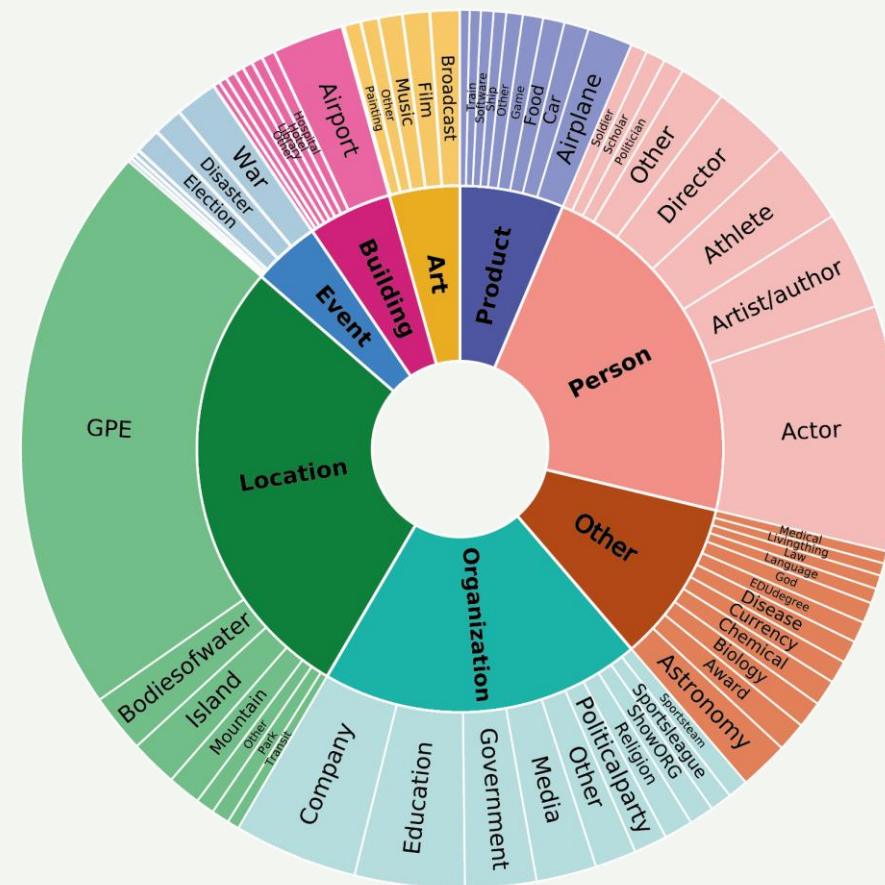
Esperimenti

Dataset benchmark

Few-Nerd



8 tipi padre + 66 tipi figlio

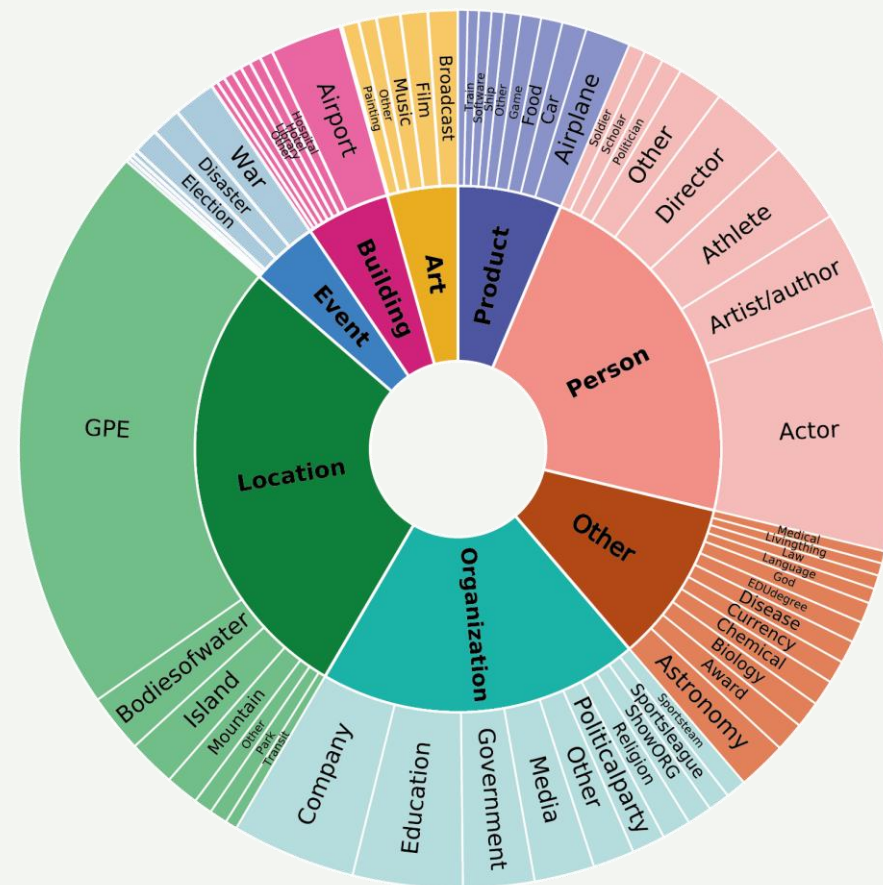
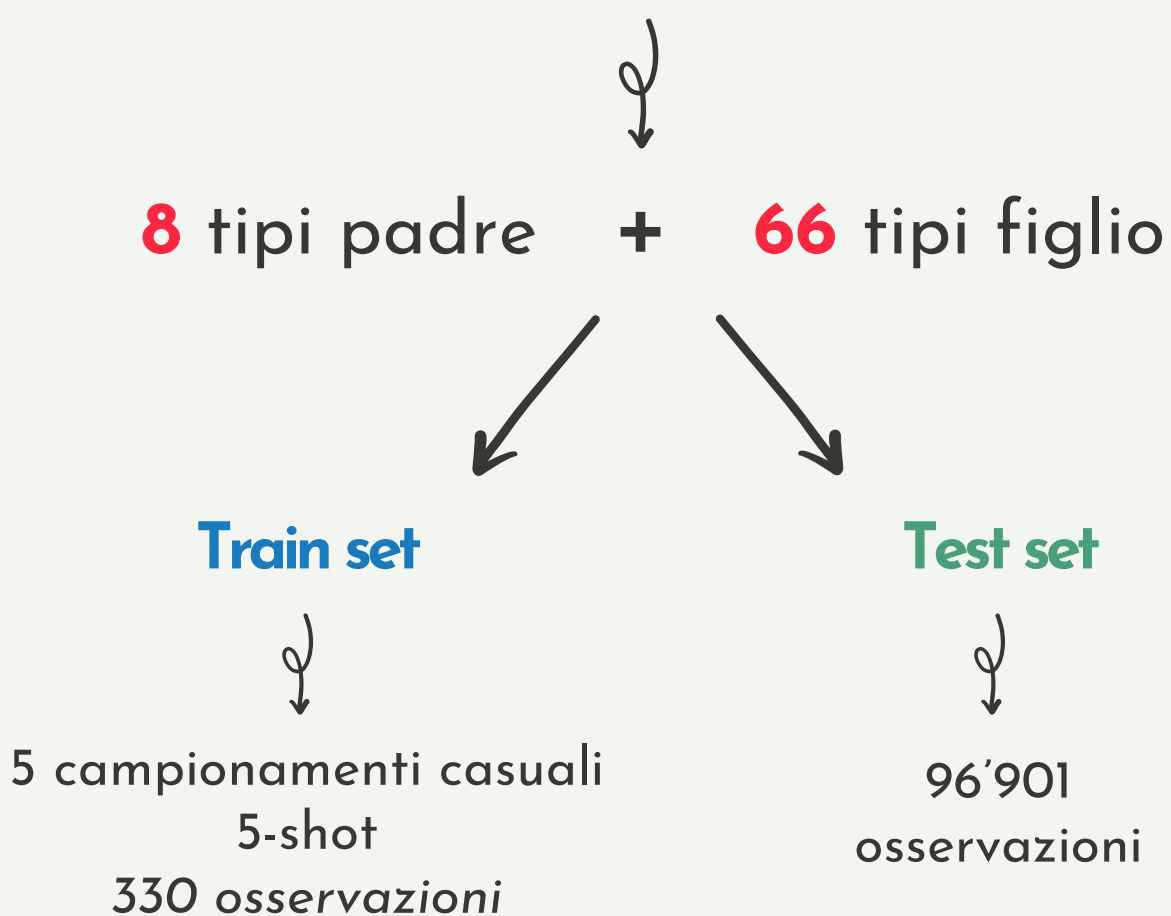


Gerarchia e distribuzione dei tipi in Few-Nerd

Esperimenti

Dataset benchmark

Few-Nerd



Gerarchia e distribuzione dei tipi in Few-Nerd

Esperimenti

Domande di Ricerca



Esperimenti

Domande di Ricerca

- 1 Come si comporta **PROMET** nella classificazione fine delle menzioni delle entità nominali nel caso **few-shot**?



Esperimenti

Domande di Ricerca

- 1** Come si comporta **PROMET** nella **classificazione** fine delle menzioni delle **entità nominali nel caso few-shot**?
- 2** Quali sono le **differenze** qualitative e quantitative **tra le due modalità implementative** di PROMET?



Esperimenti

Risultati principali

I valori riportati sono media (\pm deviazione standard) calcolata sui cinque diversi campioni casuali 5-shot

Modello stato
dell'arte

Modello	Accuracy	F1 Micro	F1 Macro (Esempi)	N° parametri	Data augmentation	MLM
ALIGNIE	57,46 ($\pm 3,20$)	68,97 ($\pm 2,93$)	68,97 ($\pm 2,93$)	3,30 mln	Sì	Sì
PROMET(F)	58,01 ($\pm 2,30$)	69,38 ($\pm 2,00$)	69,38 ($\pm 2,00$)	0,05 mln	No	No
PROMET(S)	49,40 ($\pm 1,50$)	65,36 ($\pm 1,41$)	64,60 ($\pm 1,51$)	0,06 mln	No	No

Esperimenti

Risultati principali

I valori riportati sono media (\pm deviazione standard) calcolata sui cinque diversi campioni casuali 5-shot

Modello	Accuracy	F1 Micro	F1 Macro (Esempi)	N° parametri	Data augmentation	MLM
ALIGNIE	57,46 ($\pm 3,20$)	68,97 ($\pm 2,93$)	68,97 ($\pm 2,93$)	3,30 mln	Sì	Sì
PROMET(F)	58,01 ($\pm 2,30$)	69,38 ($\pm 2,00$)	69,38 ($\pm 2,00$)	0,05 mln	No	No
PROMET(S)	49,40 ($\pm 1,50$)	65,36 ($\pm 1,41$)	64,60 ($\pm 1,51$)	0,06 mln	No	No

Modello stato dell'arte

1 PROMET(F) ha risultati in linea con lo stato dell'arte, non emergendo dai dati evidenza significativa per affermare il contrario

Esperimenti

Risultati principali

I valori riportati sono media (\pm deviazione standard) calcolata sui cinque diversi campioni casuali 5-shot

Modello	Accuracy	F1 Micro	F1 Macro (Esempi)	N° parametri	Data augmentation	MLM
ALIGNIE	57,46 ($\pm 3,20$)	68,97 ($\pm 2,93$)	68,97 ($\pm 2,93$)	3,30 mln	Sì	Sì
PROMET(F)	58,01 ($\pm 2,30$)	69,38 ($\pm 2,00$)	69,38 ($\pm 2,00$)	0,05 mln	No	No
PROMET(S)	49,40 ($\pm 1,50$)	65,36 ($\pm 1,41$)	64,60 ($\pm 1,51$)	0,06 mln	No	No

Modello stato dell'arte

1

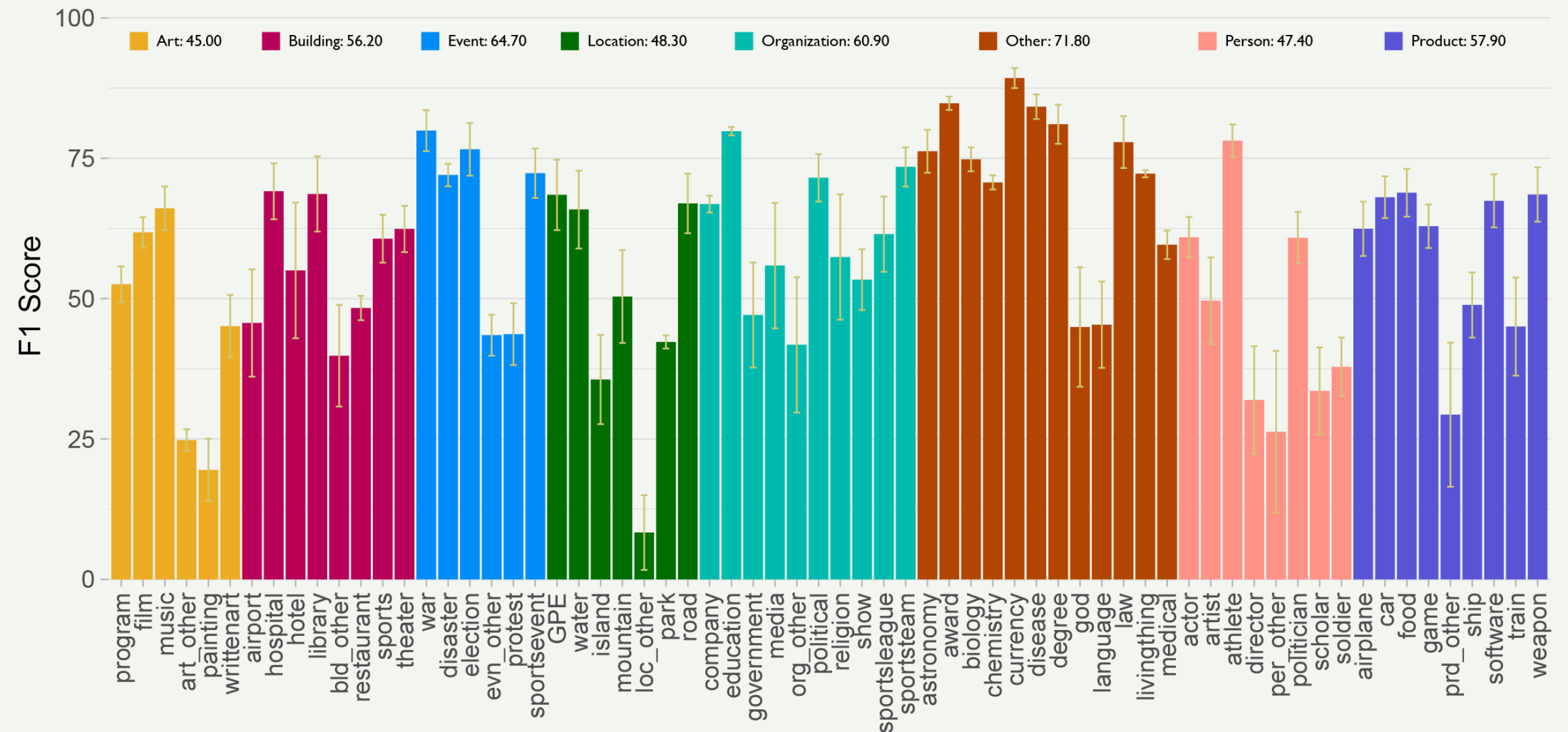
PROMET(F) ha risultati in linea con lo stato dell'arte, non emergendo dai dati evidenza significativa per affermare il contrario

2

PROMET in modalità flat ottiene in generale risultati migliori rispetto alla modalità stack

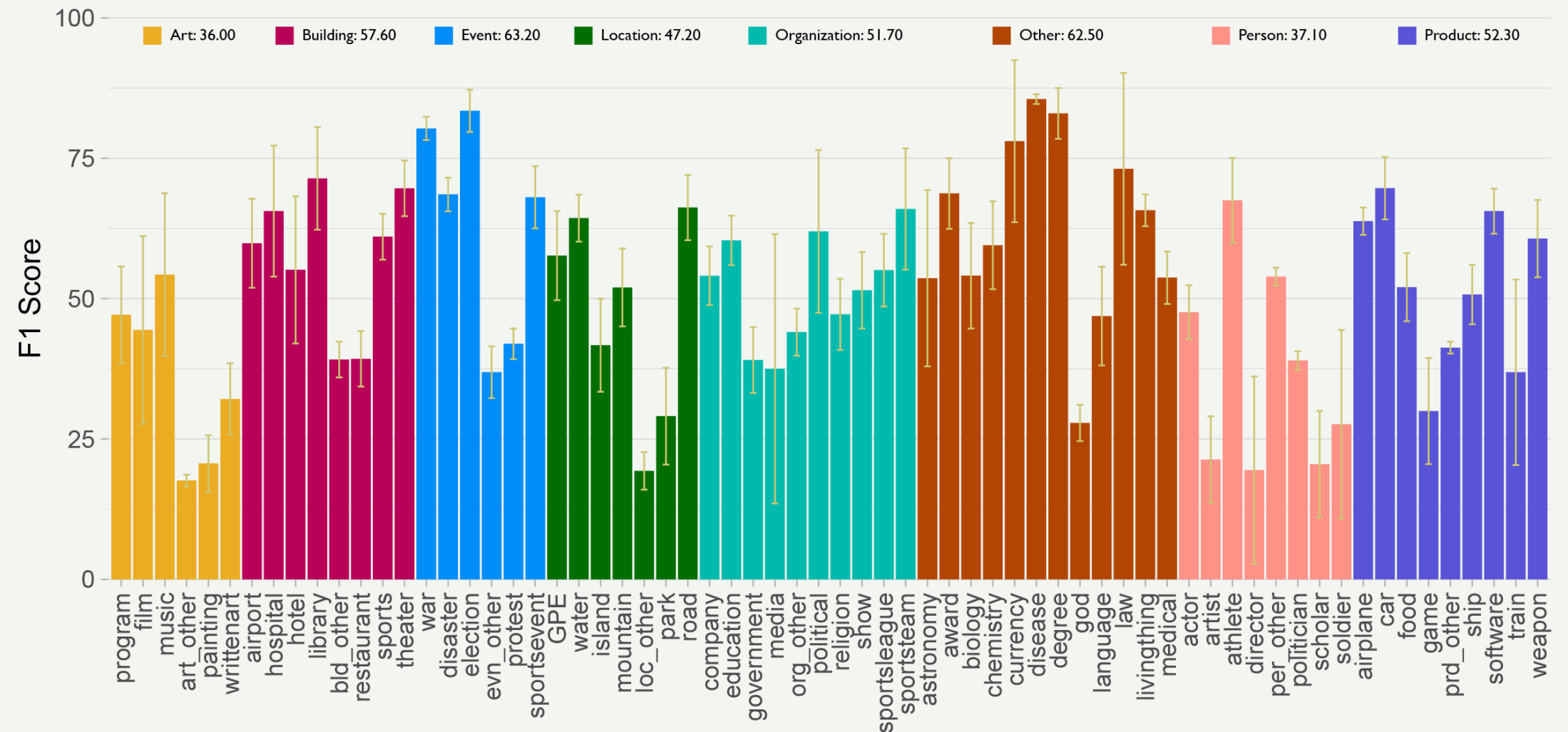
Esperimenti

PROMET(F) - Tipi Figlio



Esperimenti

PROMET(S) - Tipi Figlio



Esperimenti

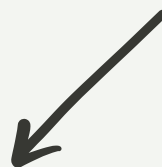
PROMET(F) vs PROMET(S)

**Perché PROMET(S) ha performance
peggiori di PROMET(F)?**

Esperimenti

PROMET(F) vs PROMET(S)

Perché PROMET(S) ha performance peggiori di PROMET(F)?



Propagazione dell'errore dal 1° livello di specializzazione al 2° livello



79,27%

Accuratezza del 1° livello di specializzazione

Esperimenti

PROMET(F) vs PROMET(S)

Perché PROMET(S) ha performance peggiori di PROMET(F)?



Propagazione dell'errore dal 1° livello di specializzazione al 2° livello



79,27%

Accuratezza del 1° livello di specializzazione

Scarsità dati d'allenamento per i classificatori del 2° livello



62,17%

Accuratezza media dei classificatori del 2° livello dato il tipo padre corretto

Esperimenti

PROMET(F) vs PROMET(S)

**Tuttavia PROMET(S) è più funzionale nel caso
d'uso di specializzazione dell'ontologia**

Esperimenti

PROMET(F) vs PROMET(S)

**Tuttavia PROMET(S) è più funzionale nel caso
d'uso di specializzazione dell'ontologia**



Grazie alla sua architettura a più livelli PROMET(S) ha una **grande flessibilità**, la quale permette al modello di **adattarsi a cambiamenti** nel set di classificazione senza dover riaddestrare tutti pesi

Esperimenti

PROMET(F) vs PROMET(S)

Tuttavia PROMET(S) è più funzionale nel caso d'uso di specializzazione dell'ontologia



Grazie alla sua architettura a più livelli PROMET(S) ha una **grande flessibilità**, la quale permette al modello di **adattarsi a cambiamenti** nel set di classificazione senza dover riaddestrare tutti pesi



Pesi del 1° livello sono indipendenti
dai classificatori del 2° livello



I classificatori del 2° livello sono tra
loro indipendenti

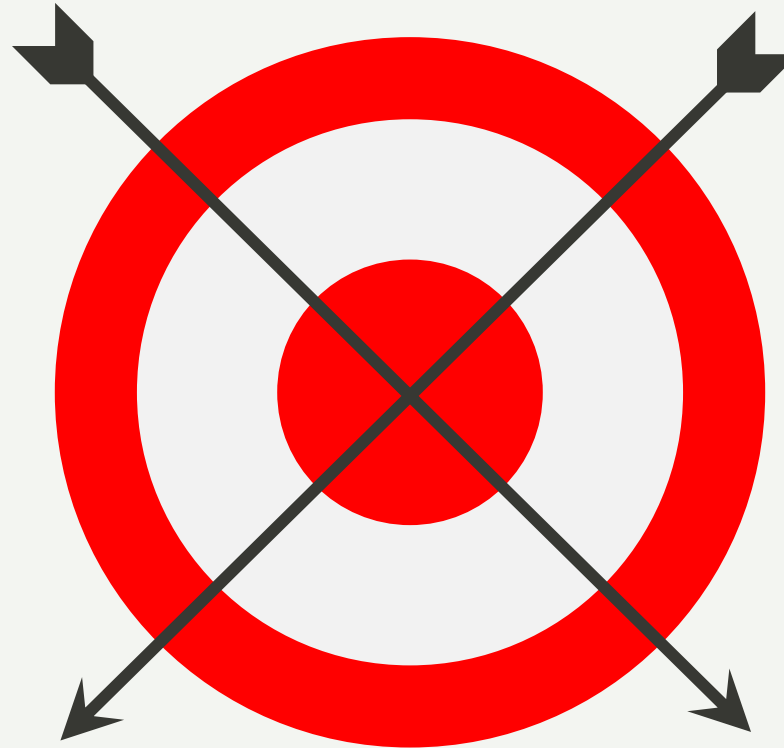
Conclusioni della Tesi

1

PROMET in modalità flat ha nello scenario few-shot risultati in linea con lo stato dell'arte

2

PROMET è più semplice rispetto ad ALIGNIE sia a livello architetturale che implementativo



3

PROMET ottiene risultati migliori in modalità flat rispetto che in modalità stack

4

PROMET in modalità stack può adattarsi a vari scenari applicativi con poco sforzo computazionale

**GRAZIE PER
L'ATTENZIONE**

