

CHIARA FAZZONE

PROFILO PROFESSIONALE

Laureata in Informatica Umanistica (percorso Tecnologie del Linguaggio), con grande interesse verso il linguaggio naturale e il Natural Language Processing (NLP). Durante l'ultimo anno, in cui ha lavorato come assegnista di ricerca su tematiche di linguistica computazionale e NLP, ha appreso e usato tecniche, modelli e librerie di Machine Learning e Deep Learning in Python.

CAPACITÀ E COMPETENZE

- Collezione, manipolazione, annotazione e analisi di dati linguistici;
- Programmazione in Python;
- Markup in XML per gestione flussi di dati;
- Jupyter notebook, Google Colab, Visual Studio Code;
- Web scraping;
- Modelli probabilistici, reti neurali e Large Language Models;
- Librerie di Machine Learning e Deep Learning, quali pytorch, tensorflow e sklearn, applicate all'NLP;
- Strumenti per l'utilizzo di Neural e Large Language Models (Hugging Face);
- Pacchetto Office;
- Standard di annotazione linguistica (Universal Dependencies);
- Conoscenza madrelingua della lingua inglese.

ESPERIENZE LAVORATIVE E PROFESSIONALI

01/2023 - Attuale

Assegnista di Ricerca

Istituto di Linguistica Computazionale "Antonio Zampolli" presso il CNR di Pisa - Pisa

Riferimento: ILC.ASS.008.2022.PI

Coinvolgimento in diverse attività di ricerca, tra cui:

- Analisi e studio della complessità testuale relativa ai contratti del settore assicurativo e bancario per la costruzione di un tool di valutazione automatica della leggibilità del testo;
- Studio delle caratteristiche linguistiche di un testo che correlano con tratti psicologici dell'autore del documento, classificazione dei testi in base a tratti psicologici dell'autore con modelli lineari e large language models (LLMs), studio del meccanismo interno dell'attenzione del LLM per la comprensione e l'interpretazione delle scelte di classificazione del modello.

10/2022 - 12/2022

Data Scientist - Stage

Wikicasa - Genova

Mantenimento e potenziamento di sistemi preesistenti, gestione flussi XML e sviluppo di nuovi modelli per l'ottimizzazione dell'estrazione e della pulizia dei dati, e per la creazione di prodotti dati per la piattaforma.

06/2021 - 02/2022

Tirocinante

Laboratorio di Linguistica Computazionale dell'Università di Pisa - Pisa

Creazione, validazione e analisi sperimentale di un dataset per l'inglese relativo alla metonimia per valutare la capacità dei Pre-Trained Language Models (es. BERT, GPT, ecc.) di catturare questa tipologia di fenomeni semantici. L'intero lavoro di costruzione del dataset e di probing di BERT è descritto nella tesi, il cui titolo è "Probing of Pre-Trained Language Models for Metonymy Classification: a new Dataset and Experiments".

ISTRUZIONE E FORMAZIONE

09/2022

Laurea Magistrale: Informatica Umanistica (LM-43) - Tecnologie del Linguaggio

Università di Pisa - Pisa

Votazione: *110 e lode* - Il percorso di studi, con i suoi corsi caratterizzanti, insegna a studiare e a trattare in modo automatico il linguaggio, dalla raccolta, l'analisi e l'annotazione dei dati fino all'addestramento di modelli di Machine Learning.

07/2018

Laurea Triennale: European Studies

University of Reading - Reading

Votazione: *110/110 e lode* - Il percorso di studi specializza lo studente nell'evoluzione, la storia, la struttura e le istituzioni dell'Unione Europea, e sulle più importanti questioni contemporanee relative al processo di integrazione europea ancora in corso. Richiede inoltre l'apprendimento di almeno una delle principali lingue europee.

TESI DI LAUREA MAGISTRALE

Titolo: "Probing of Pre-Trained Language Models r Metonymy Classification: a new Dataset and Experiments"

Relatore: Prof. Alessandro Lenci

Abstract: Il successo dei modelli pre-addestrati del linguaggio, che rappresentano lo stato dell'arte in NLP, è dovuto alle rappresentazioni vettoriali contestualizzate della lingua che sono in grado di generare. Queste, tuttavia, presentano una bassa interpretabilità. È difficile, infatti, identificare le diverse parti di informazione codificate all'interno di tali rappresentazioni, e dove e come sono codificate. L'obiettivo della tesi è di studiare le rappresentazioni estratte da BERT (Bidirectional Encoder Representations from Transformers) per capire in che misura il modello cattura informazione relativa ad un fenomeno linguistico in particolare, la metonimia.

ESPERIENZE FORMATIVE

- ♦ Partecipazione all'edizione di *Lectures on Computational Linguistics* del 2023, tenuto a Pisa dal 29 al 31 maggio.
- ♦ Completamento di *Generative AI Learning Path* tenuto da Google Cloud Skills Boost (7 luglio 2023).

PUBBLICAZIONI

Chiara Alzetta, Felice Dell'Orletta, Chiara Fazzone, Alessio Miaschi, Giulia Venturi. *Unmasking the Wordsmith: Revealing Author Identity through Reader Reviews*. In Proceedings of 9th Italian Conference on Computational Linguistics (CLiC-it), November 30 - December 2, 2023, Venice, Italy.

COMPETENZE LINGUISTICHE

♦ **Italiano:** madrelingua

♦ **Inglese:** madrelingua

♦ **Francese:** intermedio

♦ **Spagnolo:** base

PREMI

27/11/2017

Chancellor's Award - University of Reading

I vincitori del premio rappresentano l'1% degli studenti più meritevoli nell'ampio spettro delle diverse discipline proposte dall'Università di Reading. I destinatari del premio sono gli studenti che hanno ottenuto i risultati più alti nel loro corso di studio nel primo o nel secondo anno.

06/07/2018

Prize for Excellence in European Studies - University of Reading

Il premio è conferito alla laurea dal *Department of Modern Languages and European Studies* dell'Università di Reading allo studente più meritevole nel corso di studi *European Studies*.

TRATTAMENTO DEI DATI

Autorizzo il trattamento dei miei dati personali ai sensi del Decreto Legislativo 30 giugno 2003, n. 196 e del GDPR (Regolamento UE 2016/679).