

Abstract CLUB Papers 2023

Corpus Linguistics nell'era dei Social Media: le nuove sfide poste da TikTok

Melissa Donati

Alma Mater Studiorum - Università di Bologna

melissa.donati@studio.unibo.it

Paola Vernillo

Alma Mater Studiorum - Università di Bologna

paola.vernillo@unibo.it

Gli ultimi due decenni sono stati caratterizzati da un aumento esponenziale nelle interazioni web, che avvengono soprattutto attraverso l'utilizzo dei Social Media (SM), rendendo queste piattaforme fonti di dati quantitativamente e qualitativamente interessanti per ricerche linguistiche. Parallelamente, lo sviluppo di tecnologie e algoritmi per il processamento automatico del linguaggio naturale ha facilitato e velocizzato il processo di raccolta e analisi di dati linguistici dal web. Accanto alle piattaforme più diffuse e analizzate, quali Facebook, Instagram e Twitter, di recente l'attenzione è stata rivolta anche a TikTok, un social di videosharing di origine cinese che sta avendo ampio successo soprattutto tra i giovani (<35). Trattandosi di un social di nuova diffusione, tuttavia, non esiste ancora a livello metodologico una pipeline predefinita per estrarre i dati. Inoltre, la multimodalità dei contenuti che combinano testo, audio e video, pone problemi per il processamento automatico, soprattutto in fase di trascrizione. L'obiettivo di questo lavoro è quello di presentare i principali elementi di criticità relativi all'estrazione di dati linguistici da TikTok, riscontrati nel corso di uno studio volto a raccogliere e analizzare contenuti associati ai Disturbi del Comportamento Alimentare, che coinvolgono principalmente adolescenti e giovani, ossia la fascia di popolazione maggiormente attiva su questo social.