

Twita

Korpus aus 156 Mio. Italienischen Tweets von 02/2012-06/2013, teilweise mit Geo-Informationen (3,8 Mio. Tweets).

Sprache	Italienisch
Varietät	(Regional-)Standard
Sprachliche Realisierung	schriftlich
Umfang	ca. 156 Mio. Tweets
Medium	nach repräsentativen Stichwörtern gesammelte Beiträge aus dem sozialen Netzwerk Twitter
Geographischer Ursprung	Italien
Zeitliche Einordnung	02/2012-06/2013
Form der Daten	ID-Listen der Tweets, mit denen diese aufgefunden und mithilfe entsprechender Software heruntergeladen werden können. Ein Download des kompletten Korpus ist aus Datenschutzgründen nicht möglich. Zudem sind Frequenzlisten zu den häufigsten Hashtags und ein Lexikon zur Gefühlsanalyse (Sentix - Sentiment Italian Lexicon) verfügbar.
Format	txt, gzip
Annotation	tokenisiert, part-of-speech-annotiert
Quelle /Herausgeber	Valerio Basile (University of Groningen), Malvina Nissim (Università di Bologna)
Nutzungsvoraussetzungen	Zugriff frei
Link	http://valeribasile.github.io/twita/about.html
Zum Zitieren:	Basile, Valerio & Malvina Nissim. 2013. <i>TWITA. The collection of tweets from Twitter in Italian language</i> . Università degli Studi di Torino.