

Web2Corpus_IT

Das Korpus umfasst CMC-Daten (computer mediated communication) und befindet sich noch in der Aufbauphase.

Sprache	Italienisch
Sprachstufe	Standard
Sprachliche Realisierung	schriftlich
Umfang	ca. 1 Mio. Tokens
Medium	Online-Kommunikation aus unterschiedlichen Kontexten: Blogs, Foren, Chat, Social Networks
Geographischer Ursprung	Italien
Form der Daten	ODT (OpenOffice Document) oder Textdateien (.txt). Eine Excel-Tabelle mit Metadaten steht auch zur Verfügung
Format	Die Rohdaten können heruntergeladen werden sowie die annotierten Daten als XML-Dateien
Annotation	lemmatiziert, POS-Tags
Mögliche Suchabfragen	Suchabfragen sind noch nicht möglich
Quelle/Herausgeber	Isabella Chiari, Università La Sapienza di Roma
Nutzungsvoraussetzungen	Anmeldung erforderlich
Link	http://www.glottoweb.org/web2corpus/
Zum Zitieren:	Chiari, Isabella. 2011. <i>Web2Corpus_IT</i> . http://www.glottoweb.org/web2corpus/ : Università La Sapienza di Roma.