

MultiUN: Multilingual UN Parallel Text

Auswahl von Dokumenten der UN im XML-Format, verfügbar in sieben Sprachen.

Sprache	Englisch, Französisch, Spanisch, Deutsch, Arabisch, Russisch, Chinesisch
Sprachliche Realisierung	schriftlich
Umfang	ca. 300 Mio. Wörter pro Sprache
Medium	Dokumente der Vereinten Nationen, heruntergeladen von der UN-Website
Zeitliche Einordnung	2000-2009
Form der Daten	Files im XML-Format, nach einzelsprachen sortiert und zum Download verfügbar
Format	XML
Quelle /Herausgeber	Andreas Eisele, Yu Chen, DFKI (Deutsches Forschungszentrum für künstliche Intelligenz)
Nutzungs voraussetzungen	Zugang frei
Link	http://www.euromatrixplus.net/multi-un/
Literatur	Eisele, Andreas / Chen, Yu (2010): "MultiUN: A Multilingual Corpus from United Nation Documents". In: Tapias, Daniel et al. (eds.): Proceedings of the Seventh conference on International Language Resources and Evaluation. La Valletta, Malta, European Language Resources Association (ELRA), 2868-2872. Download