

MULTI-CROSSRE

A MULTI-LINGUAL MULTI-DOMAIN DATASET FOR RELATION EXTRACTION

CrossRE: manually annotated multi-domain EN dataset



	sentences				relations			
	train	dev	test	tot	train	dev	test	tot
News	164	350	400	914	175	300	369	871
Politics	101	350	400	851	502	1,616	1,831	3,949
Science	103	351	400	854	355	1,340	1,393	3,088
Music	100	350	399	849	496	1,861	2,333	4,690
Literature	100	400	416	916	397	1,539	1,591	3,527
AI	100	350	431	881	350	1,006	1,127	2,483
tot	668	2,151	2,446	5,265	2,275	7,662	8,671	18,608

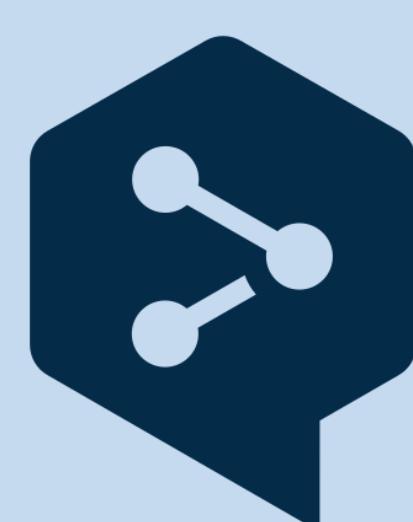
1

Cunningham played his entire 11-year
e1:person

career with the Philadelphia Eagles.
e2:organization

ROLE

Entity spans encoding
as color markup for
machine-translation



DeepL

2

In machine learning, support-vector machines [SVMs, also support-vector networks] are supervised learning models with learning algorithms that analyze data used for classification and regression analysis.

Bei maschinellen Lernen sind Support-Vektor-Maschinen [SVMs, auch Support-Vektor-Netzwerke] überwachte Lernmodelle mit Lernalgorithmen, die Daten für Klassifizierungs- und Regressionsanalysen analysieren.

In machine learning, support vector machines [SVMs, also support vector networks] are supervised learning models with learning algorithms that analyse data for classification and regression analysis.

3

Translation into
26 languages

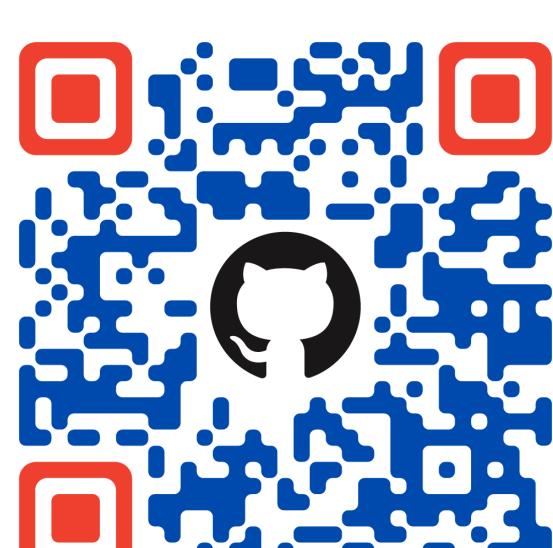


4

Manual evaluation

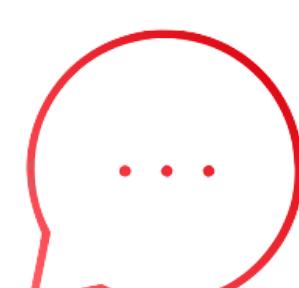
Language	UK	Finland	Italy	Germany	Denmark	Czech	Netherlands	Japan
Correct Sentence Translation	30	30	29	27	27	28	28	19
Entities Preserved	160	150	160	151	158	158	158	150
Correct Entity Translation	-	141	157	148	143	152	156	145
Correct Entity Boundaries	-	137	152	139	136	143	141	82

Elisa Bassignana
Filip Ginter
Sampo Pyysalo
Rob van der Goot
Barbara Plank



— elba@itu.dk — figint@utu.fi —

IT UNIVERSITY OF COPENHAGEN



TURKUNLP
.ORG

