Language	Task	Year	Corpora	Encoding	Docs	Topics
Bulgarian	AH Mono BG	2005 2006 2007	SEGA 2002 STANDART 2002	UTF-8	69195	149
German	AH Mono DE	2001 2002 2003	FRANKFURTER 1994 SDA 1994 SPIEGEL 1994 & 1995	ISO-8859-1	225371	155
Spanish	ROB Mono ES	2006	EFE 1994 & 1995	ISO-8859-1	454045	97
Persian	AH Mono FA	2008 2009	HAMSHAHRI 2002	UTF-8	166774	100
Finnish	AH Mono FI	2002 2003 2004	AAMULEHTI 1994 & 1995	ISO-8859-1	55344	120
French	AH Mono FR	2005 2006	LE MONDE 1994 & 1995 ATS 1994 & 1995	ISO-8859-1	177452	99
Hungarian	AH Mono HU	2005 2006 2007	MAGYAR 2002	UTF-8	49530	148
Italian	ROB Mono IT	2006	AGZ 1994 &1995 LA STAMPA 1994	ISO-8859-1 (AGZ) US-ASCII (LA STAMPA)	157558	90
Dutch	AH Mono NL	2001 2002 2003	ALGEMEEN 1994 & 1995 NRC 1994 & 1995	ISO-8859-1	190604	156
Portuguese	AH Mono PT	2005 2006	FOLHA 1994 & 1995 PUBLICO 1994 & 1995	ISO-8859-1	210734	100
Russian	AH Mono RU	2003 2004	IZVESTIA 1995	UTF-8	16716	62
Swedish	AH Mono SV	2002 2003	TT 1994 & 1995	UTF-8	142819	103