

Perché
BabelNet

BabelNet è una **rete semantica multilingue** e un'**ontologia lessicalizzata**.

BabelNet è stata creata integrando automaticamente Wikipedia (la più grande enciclopedia multilingue) con WordNet.

POS	Monosemous words	Polysemous words
Noun	22,763,265	1,134,857
Verb	6,277	5,252
Adjective	1,503	4,976
Adverb	3,748	733
Total	22,789,793	1,145,818

In alto, il numero di parole monosemiche e polisemiche in WordNet divise per part of speech.

Composition of Babel synsets: number of synonyms from the English WordNet, Wikipedia pages and translations, as well as translations of WordNet's monosemous words and SemCor's sense annotations.

	English	Catalan	French	German	Italian	Spanish	Total	
English WordNet	206,978	–	–	–	–	–	206,978	
Wikipedia	pages	2,955,552	123,101	524,897	506,892	404,153	349,375	4,863,970
	redirections	3,388,049	105,147	617,379	456,977	217,963	404,009	5,189,524
	translations	–	3,445,273	2,844,645	2,841,914	3,046,323	3,083,365	15,261,520
WordNet	monosemous	–	97,327	97,680	97,852	98,089	97,435	488,383
	SemCor	–	6,852	6,855	6,850	6,856	6,855	34,268
Total	6,550,579	3,777,700	4,091,456	3,910,485	3,773,384	3,941,039	26,044,643	

In basso il numero di sinonimi di WordNet inglese (la versione originale studiata da noi), di Wikipedia e di WordNet (che ignoriamo)

La necessità di una risorsa come BabelNet si evince da qualche analisi statistica effettuata sulla lingua inglese ed in generale su ciò che WordNet offre

Perché BabelNet

Come possiamo vedere dai dati, le risorse WordNet per lingue diverse dall'inglese sono poco sviluppate

Riuscire a districarsi nella polisemia senza risorse ricche di informazioni è sostanzialmente impossibile

D'altronde Wikipedia è una risorsa multilingua molto ricca:

Come possiamo sfruttare le informazioni contenute in Wikipedia per ottenere una risorsa che facciamo qualcosa di simile a WordNet (e anche di più) ma che al contempo tratti più lingue?

L'idea di associare manualmente le risorse Wikipedia a traduzioni di WordNet è da escludere, così come quella di costruire una WordNet per ogni lingua e poi collegarle fra loro; ricordiamo che WordNet è una risorsa annotata manualmente: pensare di ripetere il processo per ogni lingua richiederebbe dozzine di anni.