La culturonomia, un neologismo scaturito da un articolo pubblicato su Science nel 2010, è quella disciplina che studia il comportamento umano, i cultural trends ossia “ciò che le persone fanno o credono solo perché in alcune cose solo perché la maggioranza della gente crede o fa quelle stesse cose”. L’articolo "Quantitative Analysis of Culture Using Millions of Digitized Books", scritto da due ricercatori di Harvard: Jean-Baptiste Michel ed Erez Lieberman, spiegava come tramite un approccio matematico-computazionale è possibile decifrare tanto i ritmi dell’evoluzione biologica quanto quelli dei networks e del linguaggio.

È stato grazie a questi studi che Ngram Viewer (http://books.google.com/ngrams/) nel dicembre 2010 ha visto la luce: si tratta di un software, basato sul browser Bookworm, che è in grado di contare quante volte è presente un vocabolo all’interno della banca dati di Google Books, attualmente trenta milioni di testi in sette lingue diverse, di visualizzarne interattivamente i contenuti e di restituire un grafico inquadrato cronologicamente.

Ha i suoi limiti, certo, i libri finora pubblicati sono molti di più rispetto a quelli digitalizzati e gli errori provocati dalla tecnologia OCR per il riconoscimento ottico dei caratteri hanno il loro peso. Conscio di questi limiti, John Orwant, co-autore e manager del progetto, ha fatto in modo che in concomitanza dell’uscita della versione 2.0 del software, la quantità di testi fosse portata dai cinque ai trenta milioni attuali, migliorando, inoltre, sia la qualità della scansione OCR, sia la definizione dei metadati fornita dagli editori partners.

Grazie ai metadati è ora possibile effettuare delle ricerche mirate, il software è ora in grado di riconoscere il contesto in cui un argomento viene trattato in modo da non far confusione in caso di omonimie, di distinguere tra sostantivi e aggettivi, di ricercare radici o desinenze di parole.

Un altro passo verso il Web 3.0, piccolo, perché il materiale da metadatare rimane ancora tanto.

Paola Angelotti