

Tagging? Cosa "guida" gli utenti sul web, quando associano tra loro contenuti e parole chiave? Un interrogativo che interessa non solamente gli scienziati, perché risolvendolo si aprirebbe lo spazio per molte innovazioni nei servizi online. Come sistemi efficaci di protezione dallo spam (capaci di identificare le associazioni "sbagliate" di parole all'interno di email fraudolente). Oppure, servizi di pubblicità e marketing online capaci di proporre all'utente prodotti calibrati con maggior precisione. O ancora, innovazioni nei sistemi di social networking, e l'utilizzo del tagging fatto dagli utenti per catalogare enormi quantità di dati (una strada intrapresa da Google).

Un passo verso la soluzione l'hanno fatto Vittorio Loreto di INFN-CNR e Università La Sapienza di Roma, Andrea Baldassarri dell'Università La Sapienza di Roma, Ciro Cattuto della Fondazione ISI di Torino, e Alain Barrat e Gregory Schehr del CNRS francese, che hanno creato un modello di tagging capace di produrre risultati confrontabili con i dati reali estratti da siti di social networking, nello specifico del.icio.us e Bibsonomy.

Un modello semplice. L'associazione è... un problema di testa, e i processi che la regolano sono ancora piuttosto oscuri. Quello che intuitivamente sappiamo però, è che non è casuale, ma tende a muoversi lungo catene di "affinità semantica" (come "cielo" e "nuvola", "nuvola" e "bianco", e così via). Questo semplice meccanismo è stato usato dai ricercatori come base per dare vita al loro modello, che è un sistema in grado di generare post virtuali che mimano per struttura quelli reali. Centinaia di migliaia di post virtuali sono stati quindi creati e confrontati con altrettanti provenienti da siti reali, con risultati sorprendenti: i due gruppi condividono una stessa struttura complessiva, e rispondono alle stesse leggi statistiche.

Ma sorprendente. Questa grande coincidenza ha sorpreso gli scienziati. Il loro modello, volutamente generico, ha infatti ricostruito in maniera quasi perfetta la struttura dei dati reali, rispecchiandone molte importanti caratteristiche, come la co-occorrenza di tag (cioè quante volte un tag è usato assieme a un altro, si pensi a "mare" e "acqua"), o la legge che regola l'aumento del numero complessivo di tag usati dagli utenti. Un risultato brillante che i ricercatori stanno però già cercando di migliorare, perfezionando ulteriormente il modello con vari correttivi che meglio rispecchiano il comportamento degli utenti reali.