

Il database del mondo

La compilazione di un gigantesco database su tutte le vicende del mondo potrebbe aiutare i futuri prodotti di Google a comprenderci meglio.

Tom Simonite

Nonostante il suo successo, il famoso algoritmo Page Rank di Google non ha mai capito una parola dei miliardi di pagine Web a cui ha indirizzato le persone negli anni. Per questo motivo nel 2010 Google ha acquisito Metaweb, un'azienda che stava realizzando un database concepito per dare ai computer la capacità di comprendere il mondo. Due anni dopo, la tecnologia aziendale sarebbe emersa con il nome di Knowledge Graph. John Giannandrea, vicepresidente per l'ingegneria di Google e cofondatore di Metaweb, dice che questa tecnologia porterà a prodotti Google capaci di comprendere veramente le persone che li utilizzano e ciò a cui tengono.

Cos'è il Knowledge Graph?

È un distillato di tutto quello che Google sa del mondo. Un'analogia che utilizzo spesso è quella delle mappe. Per una mappa dovrete costruire un database del mondo reale e sapere che, nel mondo reale, esistono cose chiamate strade, fiumi e paesi. Si tratta quindi di creare una struttura simbolica del mondo fisico; il Knowledge Graph compie questa operazione per il mondo di idee e il raziocinio. In questo grafo della conoscenza abbiamo entità che vanno dai cibi, alle ricette, ai prodotti, alle idee filosofiche, alla storia, alle persone famose. Queste entità possono venire legate fra loro, per cui possiamo dire che due persone si sono sposate, che un determinato luogo si trova in un determinato paese, o che un film è legato a una certa persona.

Come influisce sulle ricerche via Google?

Siamo passati dal parlare di parole al parlare di quello che le parole significano.

Dalla semplice raccolta e indicizzazione dei documenti siamo ora in grado di comprendere quello di cui questi documenti parlano. Se un documento tratta di famosi giocatori di tennis, siamo in grado di apprendere che parla di sport e tennis. Ciascun pezzo di informazione che estrapoliamo, indicizziamo o ricerchiamo viene analizzato nel contesto del Knowledge Graph. Non è come comprendere completamente un testo come potremmo fare noi, ma è un passo avanti.

La capacità di comprendere quello che le persone stanno cercando aiuterà le vostre pubblicità mirate. Ma il Knowledge Graph ha altre funzioni oltre alle ricerche?

All'interno di Google il Knowledge Graph è un pezzo di infrastruttura che si sta ingrandendo ed espandendo costantemente. È uno sforzo che coinvolge tutti i rami della società. Quasi tutti i dati strutturati di tutti i nostri prodotti, come Maps, Finance, Movies e Music fanno parte del Knowledge Graph e ciò permette a tutti i responsabili di prodotto dell'azienda di essere più ambiziosi.

Google Now è un esempio di prodotto che sta cercando di dedurre lo stato in cui vi trovate per darvi consigli mirati. Una delle aree d'interesse principali del progetto consiste nella comprensione più dettagliata dell'argomento trattato. Le parole che vedete in un testo sono fondamentalmente ambigue per un computer, ma se si disponesse del Knowledge Graph e si potesse quindi comprendere come le parole siano legate fra loro, allora potrebbero venire disambiguate.

Il Knowledge Graph è completo?

Cresce di secondo in secondo. Se un'impresa locale aggiorna i suoi orari di apertura su Google, quei dati arriveranno anche sul Knowledge Graph e abbiamo algoritmi che ricercano cambiamenti in diversi siti pubblici quali Wikipedia. In sostanza prendiamo tutti questi dati grezzi e li filtriamo per stabilire il nostro livello di fiducia e la necessità o meno di aggiornare il grafo. Se una celebrità muore, per esempio, ne prendiamo atto e aggiorniamo il Knowledge Graph. Diverse persone hanno già proposto altre volte di realizzare questa sorta di rappresentazione delle nozioni comuni con un'intelligenza artificiale. Credo che Knowledge Graph sia un'implementazione molto grande e pratica di questa idea. La portata e l'accuratezza del Knowledge Graph sono probabilmente uniche nella storia.

Perché il Knowledge Graph sarebbe diverso dal Web semantico sviluppato da Tim Berners-Lee e altri?

L'idea originale del Web semantico era che le persone potessero trasmettere i dati attraverso formati standard e che un motore di ricerca come Google li avrebbe aggregati in seguito. Questa potente idea, di insegnare ai computer il mondo della conoscenza, non si stava realizzando abbastanza rapidamente e desideravamo promuoverla raccogliendo una massa critica di materiale. Riconosciamo di non disporre di tutti i dati nel mondo, ma credo che questo modello sia ugualmente utile. ■

Tom Simonite è responsabile dell'area software e hardware di MIT Technology Review USA.

