



Il data center di Facebook in Svezia

SIMON DAWSON (BLOOMBERG VIA GETTY IMAGES)

I server non garantiscono una memoria di ferro

Adrian Lobe, Neue Zürcher Zeitung, Svizzera

Siamo circondati da milioni di dati, quindi da enormi conoscenze. Ma dobbiamo capire come conservarli, perché i vecchi formati potrebbero non essere più leggibili in futuro

Nel 2010 l'allora amministratore delegato di Google Erich Schmidt disse che in due giorni producevamo una quantità d'informazioni pari a quella che l'umanità aveva accumulato dagli albori della sua storia fino al 2003. Schmidt stimava che questa massa d'informazioni corrispondesse almeno a cinque exabyte, più o meno

l'equivalente di 560 libri per ogni persona nel mondo. Ma quelle cifre sono invecchiate: nel 2017 sono state spedite ogni minuto 527 mila foto con Snapchat e si sono fatte 3,6 milioni di ricerche su Google.

Ogni giorno su Twitter vengono generati circa 12 milioni di petabyte, ovvero 2.300 volte la quantità di dati conservati nella biblioteca del congresso americano, che possiede un patrimonio di 38,8 milioni di libri, 70 milioni di manoscritti e 14,2 milioni di fotografie.

Tweet poco meritevoli

Come si capisce chiaramente da questi numeri, i famosi *big data* sono effettivamente grandi, e l'era digitale ha già raggiunto dimensioni inaudite. Certo, i dati non sono

necessariamente informazioni: di per sé non dicono nulla, bisogna fargli domande specifiche. Solo attraverso la creazione di strutture narrative i dati diventano informazioni. La questione ora è come conservare nel tempo queste informazioni che custodiscono anche il sapere sul funzionamento della nostra società.

Stando alla legge di Moore, la capacità di calcolo – così come la memoria dei microprocessori – raddoppia all'incirca ogni diciotto mesi. Questa legge si è dimostrata valida per cinquant'anni, ora però sembra meno attendibile. Nemmeno se raddoppiasse ogni sei mesi la capacità di memoria basterebbe a salvare i dati (in crescita esponenziale) dell'internet delle cose, che vede interconnessi tutti i nostri oggetti, dalle auto agli spazzolini da denti. Tanto più che la longevità dei supporti di memoria si riduce costantemente.

Pensiamo ai cd rom o ai vhs, che sono decisamente più deperibili dei libri: dopo venti o trent'anni non sono più leggibili. I ricordi delle vacanze fatte da bambini sono già irrimediabilmente persi.

Da poco la biblioteca del congresso ha

rinunciato al suo progetto più ambizioso: archiviare tutti i tweet pubblicati. Ha deciso di selezionare i tweet da conservare, come fa con gli altri documenti. Dal 2000 a oggi ha salvato 525 terabyte di materiali presi dalla rete, ma il solo volume di dati dei tweet rischia di sovrappassare la capacità di archiviazione della biblioteca.

Tra i tweet da conservare ci sono quelli del presidente americano Donald Trump: 37mila tweet che riempiono da soli un intero volume della biblioteca. Ma i messaggi di Trump, spesso molto poco presidenziali e a volte decisamente offensivi, meritano sul serio di essere tramandati?

Le biblioteche sono luoghi dove si affrontano le questioni fondamentali della nostra società. Tra queste c'è la questione dell'eredità culturale, e dei modi in cui va preservata e inserita nel canone del sapere. Includere i tweet diffamatori di Trump nel patrimonio digitale della biblioteca del congresso insieme ad altri milioni di commenti carichi d'odio significa nobilitarli e quindi legittimarli socialmente? O indipendentemente dalla loro qualità e dal loro tono, sono testimonianze dello straordinario consolidamento del messaggio politico sintetizzato in 140 caratteri?

Pergamena digitale

Internet Archive, un'organizzazione non profit, si è data l'obiettivo di archiviare siti web, testi, foto, file audio e video, e di creare una sorta di biblioteca di Alessandria digitale. La memoria di internet riposa in una vecchia chiesa di San Francisco: 310 miliardi di immagini provenienti da siti internet e catalogate, tra cui anche documenti storici, come la pagina web del New York Times dell'11 settembre 2001. Il progetto si basa sulla convinzione che la capacità di memoria di internet sia fragile e che informazioni importanti potrebbero andare perdute.

Mettiamo di trovarci nel 2100 e di fare una ricerca su Google. Sarebbe ancora possibile rintracciare siti indicizzati nel 2018? E soprattutto, esisterebbero ancora i motori di ricerca? L'informatico statunitense Vint Cerf, considerato uno dei padri di internet, già nel 2015 metteva in guardia su una possibile "età buia" digitale.

Le prossime generazioni rischiano di soffrire di una sorta di amnesia digitale perché i vecchi formati potrebbero non essere più leggibili. "Possiamo creare grandi archivi di contenuti digitali ma nel tempo

Come potremo far capire i fluidi algoritmi, le leggi della nostra epoca, alle prossime venti o trenta generazioni?



potremmo non essere più in grado di sapere cosa contengano", diceva Cerf. Per questo ha proposto la stesura di una pergamena digitale su cui trascrivere e conservare ogni singolo software. È un'idea condivisibile, perché l'archiviazione non è fine a se stessa. Per lo studioso di letteratura e mezzi di comunicazione Roberto Simanowski, il vero problema di internet non è la mancanza di memoria, ma "una capacità di dimenticare attenuata". Ogni comunicazione quotidiana viene salvata da qualche parte e va a riempire quelle riserve di dati a cui attingono le società di marketing e i servizi segreti per estrapolare informazioni sulla società.

La crescente funzione di memoria di internet rappresenterebbe la "radicalizzazione tecnica" della mania dell'archiviazione e dell'esplosione dei discorsi sull'argomento. Per quanto suoni paradossale, più dati abbiamo più cresce l'incertezza sulla loro origine. Abbiamo sempre più informazioni, ma sappiamo sempre meno sulle informazioni in sé.

Come differenziare i dati spazzatura? Come consolidare le conoscenze della nostra epoca considerando la fragilità e la breve vita dei supporti di memoria? Quale sistema di documentazione svilupperemo? Quanti byte servono per rendere un evento significativo e quindi degno di essere conservato? Come verranno riportate narrazioni alternative? E chi decide sulla cultura della memoria: chi ha i server più capienti?

Cosa succederebbe se Mark Zuckerberg dicesse: "Le pagine su Facebook dei nostri due miliardi di utenti sono un'eredità importante e meritano di essere tramandate

al posto degli archivi dei quotidiani"?

Facebook e Google sono da tempo un archivio della nostra civiltà. Se nel medioevo erano gli ecclesiastici e i principi a controllare i luoghi che custodivano la memoria, ora quel ruolo è svolto dagli algoritmi delle aziende tecnologiche.

L'informatico Clifford Lynch ha evidenziato le difficoltà della documentazione nell'era digitale. Da un lato ci sono i "testimoni robot" o le "popolazioni sintetiche", come gli eserciti di bot. Dall'altro i sistemi di algoritmi strutturati in una moltitudine di sottosistemi, dai consigli automatici negli acquisti online fino alla polizia predittiva. Sistemi centrali, le cui operazioni di calcolo non sono documentate da nessuna parte.

Gli algoritmi sono scatole nere. Con i suoi 3800 anni d'età il codice di Hammurabi, proveniente dall'antica Mesopotamia, è una delle più antiche raccolte di leggi scritte ed è conservato nel museo del Louvre a Parigi inciso nella pietra. Ma come potremo far capire i fluidi algoritmi, le leggi della nostra epoca, alle prossime venti o trenta generazioni?

Cancellare tutto

Nella storia il sapere è stato più volte rigettato o distrutto. Le biblioteche sono state date alle fiamme o svuotate dei loro volumi, i libri sono stati bruciati, i pensatori "eretici" cancellati dalla storia.

Anche oggi le biblioteche selezionano il sapere ed eliminano le testimonianze che per loro sono trascurabili. La grande differenza tra i supporti analogici e quelli digitali è che anche in quelli analogici gli scarti possono essere recuperati.

Gli archeologi possono leggere il passato nelle discariche della storia e riesumare proprio quegli strati che le culture precedenti non consideravano degni di essere conservati. Nell'era digitale invece, i dati e a volte intere identità possono essere cancellati con un semplice clic, se non c'è una copia di riserva. E non è da escludere che, diffondendo virus informatici, gli hacker possano annientare intere testimonianze della nostra epoca.

Il collettivo Anonymous ha lanciato il grido di battaglia: "Noi siamo legione. Noi non dimentichiamo. Noi non perdoniamo". Ironia della storia, è proprio la generazione ossessionata dalla memoria, e che fa del non dimenticare un'ideologia, a far temere l'oblio collettivo. ♦ nv