
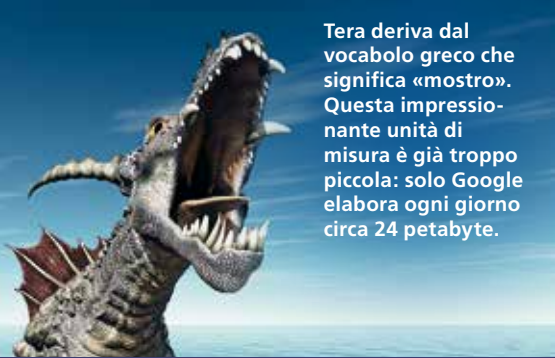



Nel 1956 IBM immetteva sul mercato il primo disco rigido ad uso commerciale, in grado di memorizzare poco più di 4 megabyte.



Oggi la più grande chiavetta USB in commercio ha una capacità di 2 terabyte, cioè 2 miliardi (milioni di milioni) di megabyte.




Tera deriva dal vocabolo greco che significa «mostro». Questa impressionante unità di misura è già troppo piccola: solo Google elabora ogni giorno circa 24 petabyte.



È stato stimato che nel cervello umano entrino 2,5 petabyte di dati.

Mega, giga, tera, ...

1 bit	(la più piccola unità di misura digitale)
1 byte	= 8 bit
1 kilobyte	= 1000 byte
1 megabyte	= 1000 ² byte o 1000 kilobyte
1 gigabyte	= 1000 ³ byte o 1000 megabyte
1 terabyte	= 1000 ⁴ byte o 1000 gigabyte
1 petabyte	= 1000 ⁵ byte o 1000 terabyte
1 exabyte	= 1000 ⁶ byte o 1000 petabyte
1 zettabyte	= 1000 ⁷ byte o 1000 exabyte
1 yottabyte	= 1000 ⁸ byte o 1000 zettabyte



Entro il 2025 la quantità di dati immagazzinati nel mondo registrerà una crescita esponenziale arrivando a 163 zettabyte. Circa dieci volte i dati del 2016.

Colophon

SATW Technoscope 01/20 | Febbraio 2020
www.satw.ch/technoscope
www.satw.ch/technoscope
Idea e redazione: Beatrice Huber | Ester Elices
Collaboratori di redazione: Christine D'Anna-Huber | Alexandra Rosakis
Grafica: Andy Braun
Foto: Adobe Stock
Foto di copertina: Adobe Stock
Traduzione: Ars Linguae
Stampa: Egger AG

Abbonamento gratuito e ordini supplementari

SATW | St. Annagasse 18 | CH-8001 Zurigo
technoscope@satw.ch | Tel +41 44 226 50 11
Technoscope 2/20 uscirà a maggio 2020 sul tema «Food»

Big Data

Come vengono trattati i nostri dati

Big Data

Estrapolare conoscenze dai dati

Smartphone, carte di credito, GPS o fitness tracker: ovunque andiamo, lasciamo dietro di noi una scia di dati. Ma non solo noi esseri umani: sono sempre di più i sensori all'interno di dispositivi e apparecchi che dipendono da Internet e producono dati. Nelle auto a guida autonoma si producono fino a 4 terabyte al giorno. Un terabyte è il numero 1 seguito da 12 zeri. Una cifra inimmaginabile per la sua grandezza.

Big Data significa da un lato che vengono generate quantità sempre maggiori di dati dalle fonti più disparate. E dall'altro che oggi è possibile conservarli in maniera molto più efficiente e conveniente rispetto al passato – su dischi rigidi e memorie flash, e nel cloud. Secondo gli esperti in futuro la quantità di dati memorizzati raddoppierà ogni tre anni.

Conoscenze preziose grazie ai dati

La particolarità dei Big Data non è tanto questa enorme mole di dati di per sé, quanto piuttosto la possibilità di ricavarne conoscenze preziose grazie a calcolatori sempre più potenti. La tecnologia ha per così dire «dopato» l'elaborazione dei dati: software speciali e algoritmi ad apprendimento automatico ripuliscono in un attimo gli enormi quantitativi di dati memorizzati alla ricerca di modelli, correlazioni e pattern ripetitivi nascosti. Un processo che consente di trarre conclusioni e rilevare tendenze in molti settori – a partire dalla ricerca e dalla medicina per arrivare alla finanza, al marketing e anche all'agricoltura: quali sono le aree a rischio di perdita del raccolto? Quali pazienti rispondono meglio a un farmaco? Quale gruppo di clienti potrebbe essere interessato al nuovo prodotto? Come migliorare l'efficienza nella gestione dei flussi di traffico?

Ogni giorno nel mondo viaggiano **294** miliardi di e-mail, a cui si aggiungono **5** miliardi di ricerche, **65** miliardi di messaggi su WhatsApp e **500** milioni di tweet.

Da dove provengono tutti i dati per i Big Data?



Rilevamento di tracce

Ogni attività su Internet viene registrata: che video hai visto su YouTube, che prodotti acquisti, che musica preferisci, quanti «like» hai su Instagram, chi sono i tuoi amici, quanto sei in forma (grazie allo smartwatch), che percorso hai fatto (grazie alla bici a noleggio). Da qui vengono estrapolati dati che magari tu non hai mai indicato, ad esempio la tua età, il tuo grado di istruzione, i tuoi hobby, ma che valgono oro per tante imprese: infatti, grazie a questi dati personali possono proporre servizi e pubblicità mirate.



Commercio di dati

Esiste un fiorente commercio di dati, che vengono venduti e acquistati. Per molti utenti non è un problema perché «non hanno niente da nascondere». Ma qualcuno può anche fare un cattivo uso dei nostri dati, oppure possiamo subire un danno a causa dei nostri dati – ad esempio l'aumento del premio della cassa malati dovuto ad abitudini poco sane – o possiamo addirittura diventare vittima di un furto di identità. In Cina la sorveglianza da parte dello stato, così come la punizione o la ricompensa dei cittadini sulla base dei dati personali, è già realtà. In Svizzera la protezione dei dati è regolata da una legge federale, attualmente in fase di revisione per rispondere alle nuove realtà su Internet.



Le potenzialità degli Open Data

Ma i Big Data sono molto di più che dati personali. I cosiddetti Open Data, i dati aperti, sono liberamente accessibili a tutti ed elaborabili da chiunque. Si tratta ad esempio di portali di conoscenze, come Wikipedia, o di statistiche, come quelle pubblicate dall'Ufficio federale di statistica, informazioni sul traffico, dati su eventi ambientali attuali, ma non solo, anche software e opere d'arte come immagini o video. Un esempio? La Piattaforma Open Data dei TP svizzeri ((Link für PDF: <https://opentransportdata.swiss/de/>)), che mette a disposizione dati su orari e fermate per consentire lo sviluppo di app o altri prodotti, o la creazione di statistiche.



I Big Data spiegati in 3 minuti:

https://www.youtube.com/watch?v=uH813u7_b0s

Dove mi trovo?



©Verkehrsbetriebe Zürich

Immagina di dover cercare il mezzo di trasporto pubblico più veloce per arrivare in tempo a un appuntamento e di avere in mano solo l'orario ufficiale della Svizzera in formato cartaceo – una vera sfida.

Per fortuna, da quando è stato pubblicato il primo orario ufficiale nel 1905, le cose sono molto cambiate. Diverse app si fanno carico della fastidiosa ricerca del collegamento migliore, ma non si limitano a questo, vendono anche il biglietto e forniscono informazioni aggiornate su eventuali ritardi o deviazioni.

Miglior flusso di traffico grazie alle connessioni in tempo reale

La mobilità del futuro, come ad esempio quella pianificata dalla città di Zurigo, intende andare oltre e connettere tra loro tutti gli utenti della strada – rete di trasporto pubblica, car sharing, taxi, biciclette, bici a noleggio – in tempo reale www.stadt-zuerich.ch/vbz/index/mobilitaet-der-zukunft. Anche le auto private dovranno comunicare tra loro e con l'infrastruttura di traffico. Il che, tra gli altri vantaggi, dovrebbe anche contribuire a migliorare il flusso di traffico. Grazie a Google Maps è già possibile informarsi sugli ingorghi in tempo reale: Google rile-

va la posizione individuata tramite il GPS degli smartphone e, in base al loro numero e alla loro localizzazione, riproduce sulla mappa la situazione del traffico.

Dati come segnali stradali

E a proposito di mappe: dobbiamo ringraziare proprio i Big Data se oggi riusciamo a trovare così facilmente il tragitto da A a B. Google ad esempio lavora con innumerevoli partner che forniscono coordinate dettagliate per rappresentare la situazione geografica precisa e aggiornata. Street View inoltre fornisce i dati per migliorare la mappa leggendo i cartelli stradali raffigurati e allineandoli con quanto riportato sulla mappa. L'impiego dei dati satellitari aiuta a scoprire eventuali variazioni geologiche o strutturali.

Interessante per tutti noi è anche il fatto che noi stessi possiamo contribuire al valore d'uso della mappa nel momento in cui trasmettiamo foto o valutazioni. Gli algoritmi, così come il personale in carne e ossa, gestiscono, confrontano e interconnettono questi set di dati stratificati migliorando continuamente le mappe.

La mobilità del futuro

I dati contribuiscono a migliorare il flusso di traffico.

Big Data by you



grafando o perfino giocando a un videogame. Attraverso l'app allyscience.ch ad esempio chi soffre di febbre da fieno può documentare i suoi disturbi e quindi contribuire allo sviluppo dei futuri sistemi di segnalazione precoce e delle terapie per l'allergia ai pollini.

Sono tante le questioni scientifiche che necessitano di un'enorme quantità di dati faticosamente raccolti. Non si tratta necessariamente di parametri misurati in laboratorio con sistemi complicati, ma di dati che possiamo osservare tutti noi nel nostro ambiente. La Citizen Science – il processo in cui i cittadini sono contemporaneamente protagonisti e destinatari della ricerca – coinvolge gli appassionati di scienze allo scopo di rilevare velocemente tali dati.

Tutti aiutano tutti

Di solito i ricercatori avviano e gestiscono i progetti, valutano i dati e li rendono accessibili alla collettività, che, a seconda del progetto, può venire coinvolta nel rilevamento dei dati o anche già in fase di pianificazione dei progetti e nella successiva analisi dei risultati. In quest'ottica l'Università di Zurigo e il Politecnico federale di Zurigo hanno creato il centro di competenze Citizen Science <https://citizenscience.ch/>. La gamma degli argomenti di ricerca che possono essere studiati con l'aiuto di volontari è molto ampia. Sul sito www.schweiz-forscht.ch vengono presentati i diversi progetti, i quali spaziano dalla diffusione della salamandra alpina alla varietà dei dialetti svizzeri fino alla correzione degli errori quantistici. Tutti gli interessati possono iscriversi e prendere parte alla ricerca osservando, raccogliendo, foto-

La qualità dei dati deve essere una priorità

Una delle difficoltà di questa raccolta variabile di dati è rappresentata dal controllo di qualità. Può capitare, soprattutto nel caso in cui il rilevamento dei dati si riveli estremamente complesso, che i dati vengano involontariamente sfalsati. A prescindere dal controllo diretto da parte della community, i dati quindi vengono anche verificati mediante algoritmi e filtrati allo scopo di eliminare gli inserimenti errati.

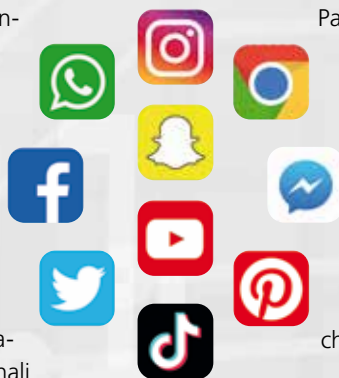


Tutto uguale dentro la bolla

Quali sono le fake news e quali invece le notizie vere? Una volta i quotidiani più autorevoli ci aiutavano a districarci nella giungla di informazioni, spiegavano e classificavano le notizie. Nell'era di Internet la giungla di informazioni è diventata ancora più fitta, le fonti ancora più numerose, la loro origine e la loro veridicità ancora meno trasparenti. Nel frattempo i tradizionali organi di informazione hanno perso importanza abdicando di fatto al loro ruolo di «guardiani». Sono soprattutto i giovani ad aver voltato loro le spalle. Ragazzi che preferiscono informarsi attraverso aggregatori di notizie come Reddit, blog e fori online. Secondo lo studio James 2018, circa il 51% dei giovani svizzeri usa i social network come fonte di informazione quotidiana.

News personalizzate

Gli aggregatori di notizie si basano sugli algoritmi, che fanno ciò che prima spettava alla redazione: scelgono le informazioni da presentare ai lettori. Ma ai tempi dei Big Data le notizie vengono scelte selettivamente, su misura per i singoli clienti.



Partendo da ogni sorta di dati, quali la provenienza, l'età e il sesso, combinati ai risultati delle ricerche effettuate, ai siti visitati e al comportamento di navigazione dell'utente, gli algoritmi tracciano un profilo della personalità dal quale ricavano la tipologia di notizie che potrebbero interessarlo o divertirlo di più. Tutte le altre non gli vengono nemmeno proposte. Il che alla lunga gli dà l'impressione che il mondo sia formato da persone che condividono le sue stesse idee. Chi non si confronta mai con opinioni diverse dalle proprie o non viene stimolato a riflettere e a imparare di più, è rinchiuso nella bolla di filtraggio.

Fuori dalla bolla

Il migliore antidoto è la varietà: scoprire come lo stesso argomento viene trattato altrove – magari, tanto per cambiare, anche nei media tradizionali.

Vedere chi c'è dietro una notizia, la serietà della fonte e a chi si rivolge la notizia. E poi formarsi la propria opinione.

Cioccolato, premi Nobel e conclusioni errate

Gli algoritmi intelligenti possono filtrare e ricavare informazioni preziose da diversi set di dati. Ad esempio, perché la Svizzera vanta un numero impressionante di Premi Nobel. Sorprendentemente, la causa andrebbe ricercata nell'elevato consumo pro capite di cioccolato, superiore a qualunque altra nazione. Questo, perlomeno, è ciò che sembra comprovare il grafico sottostante. O magari non è così?



Prendiamo un altro esempio: dalla seconda guerra mondiale in Svizzera è diminuito sia il numero di cicogne che il numero di figli per nucleo familiare. Questo prova forse che sono proprio le cicogne a portare i bambini – come si credeva un tempo? No di certo. Il fatto che oggi i nostri tetti ospitano meno cicogne è triste, ma non è decisamente la causa del calo delle nascite.

Lo studente di Harvard Tyler Vigen ha scovato una serie di curiose correlazioni apparenti. Molte fanno sorridere, ma evidenziano anche un problema che chi si occupa di statistica conosce

bene. Ovvero: «correlazione non significa causalità». Il nesso tra due cose non significa affatto che una provochi l'altra.

In informatica esiste un detto: «garbage in, garbage out». Tradotto letteralmente in italiano significa: «spazzatura dentro, spazzatura fuori». Riferita ai Big Data, la frase intende dire che la forza espressiva di un risultato prodotto da un algoritmo non dipende solo dai dati immessi, ma anche dal fatto che l'algoritmo sia stato programmato in maniera tale da calcolare i parametri corretti.



Per ulteriori informazioni:

<https://www.giovanimedia.ch/it/temi/fake-news-manipolazione.html>



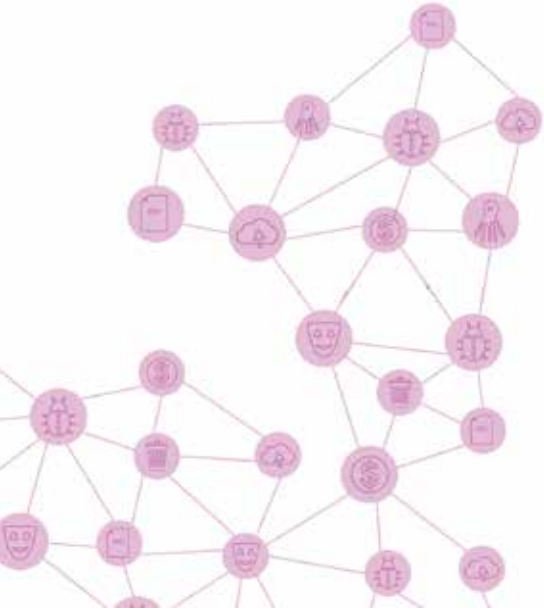
Curiose correlazioni apparenti:

<http://www.tylervigen.com/spurious-correlations>



Big Data per le aule scolastiche

I Big Data sono un tema complesso ma troppo importante per permettersi il lusso di non comprenderlo. Il Programma nazionale di ricerca PNR 75 «Big Data» e il Museo delle Comunicazioni hanno unito le forze per sviluppare uno strumento didattico in materia di Big Data rivolto agli studenti di grado secondario I e II.



Grazie a questo strumento, suddiviso in più parti e disponibile in tedesco e in francese, l'informatica diventa comprensibile. L'accento è posto sul mondo dei ragazzi. Gli studenti imparano il significato dei Big Data e a confrontarsi criticamente con questo tema, nonché a maturare una consapevolezza delle conseguenze a livello personale del fenomeno. Perché in fondo anche i ragazzi non sono solo semplici consumatori. Spetta a ognuno di noi il compito di plasmare il futuro.



Strumento didattico «Big Data» (in francese): www.mfk.ch/fr/bigdata

Esprimi il tuo talento!



#SwissTecLadies

swiss **TecLadies**
by satw



«Big Data – Il gioco»

A titolo introduttivo, nella prima parte viene privilegiato il metodo ludico. L'aspetto giocoso consiste nel risolvere un enigma, sulla falsariga di una black story. Una ragazza americana rimane involontariamente incinta. Non lo dice a nessuno e cerca di comportarsi senza dare nell'occhio. Neanche il padre sa niente. Nel frattempo la catena di supermercati «Target» spedisce alla ragazza alcuni buoni per l'acquisto di abiti premaman e articoli per bebè. Com'è possibile? All'inizio delle due ore di lezione, agli studenti viene posta una domanda chiave in termini enigmatici. Nel corso del gioco i ragazzi raccolgono le informazioni che possono aiutarli a rispondere alla domanda chiave e a risolvere tutti insieme il mistero.

«Big Data – Il laboratorio»

Nella seconda parte vengono approfonditi sette aspetti dei «Big Data». A seconda del livello e delle conoscenze – ma anche delle esigenze – della classe, vengono scelti gli argomenti tra i sette aspetti trattati.

- «Perché segreti?» – Sfera privata
- «Niente è gratis!» – Dati preziosi
- «Cosa riceviamo da vedere?» – Bolla di filtraggio
- «Sanno esattamente cosa facciamo...» –

Sorveglianza totale

- «Come funziona la prevenzione in rete?»
- Protezione dei dati
- «Salvare il mondo con i Big Data?» – Potenzialità dei Big Data
- «Ridatateci i nostri dati!» – Open Data

«Big Data – Il Trail»

L'ultima parte è rappresentata da un gioco pedagogico nel Museo delle Comunicazioni di Berna. A livello di contenuti, «Il Trail» integra il capitolo de «Il Laboratorio», ma ci si può giocare anche senza aver precedentemente elaborato l'argomento durante la lezione. «Il Trail» si svolge nel futuro, precisamente nel 2080. Gli uomini hanno ceduto il controllo totale dei propri dati al grande gruppo Amathron. Ma la base clandestina si oppone a questo status quo e vuole riprendersi il controllo dei dati. I giocatori si trovano in un simulatore di addestramento della base clandestina e compiono tutte le sessioni di addestramento per prepararsi ad affrontare tutte le sfide e le avventure nel mondo digitale. Alla fine delle sessioni di addestramento i giocatori iniziano la loro prima missione come attivisti. Vengono catapultati nel passato – la nostra epoca – per poter esercitare la loro influenza sulle decisioni strategiche del nostro tempo.

Fare ricerca con i Big Data

Ma il lavoro non finisce con la raccolta di grandi quantità di dati. Seguono le fasi di memorizzazione, gestione, elaborazione, valutazione, rappresentazione e interconnessione dei dati: tutti compiti facilmente gestibili per piccoli set di dati, ma che invece comportano una sfida gigantesca per i Big Data. È necessaria un'attività di ricerca, come ad esempio nel Programma nazionale di ricerca (PNR) 75 «Big Data». Nel PNR vengono attuati i progetti di ricerca che contribuiscono a risolvere importanti problematiche attuali, o addirittura a vincere le grandi sfide che attendono la Svizzera. Il PNR 75 analizza e studia a fondo anche i campi di ricerca sottostanti.

Uno dei punti cruciali dei Big Data è il tempo impiegato dagli algoritmi per l'elaborazione dei dati. In particolare per quanto riguarda il settore dell'apprendimento automatico si punta a sviluppare algoritmi più potenti e più veloci.

http://bit.ly/nfp75_algorithmen

Un'ulteriore sfida riguarda la memorizzazione corretta e di facile applicazione di enormi quantità di dati. Uno degli approcci proposti per ridurre la necessità di memoria è l'analisi dei dati in tempo reale: i dati in ingresso non vengono solo memorizzati ma vengono direttamente analizzati. Questi sistemi di elaborazione del flusso di dati potrebbero essere utili anche agli esperti di altre discipline oltre all'informatica.

http://bit.ly/nfp75_Datenstromanalytik

Gli inserimenti nelle banche dati possono occupare centinaia di colonne, cosa non prevista negli attuali linguaggi di programmazione. Ciò rende più difficile l'utilizzo delle banche dati stesse. Ecco perché un altro progetto di ricerca mira a migliorare l'interazione tra i diversi linguaggi di programmazione e le banche dati.

http://bit.ly/nfp75_Datenstromanalytik

I Big Data di per sé rappresentano un'enorme area di ricerca in cui sono tante le questioni ancora irrisolte. Affinché i Big Data possano esprimere tutto il loro potenziale, non è necessaria solo la figura dello scienziato dei dati, ma è indispensabile coinvolgere anche gli esperti di altre discipline.

Scelta degli studi e del lavoro



Giovanni Lombardi, psicologo orientatore scolastico e professionale, Ufficio dell'orientamento Breganzona

Mi sono iscritto a un canale YouTube dedicato alle trasformazioni digitali e al mondo delle nuove tecnologie. Nell'ultimo video si parlava della rivoluzione dei Big Data e questo tema ha suscitato in me un grande interesse. Se un giorno volessi diventare uno scienziato dei dati, quale percorso mi converrebbe intraprendere?

(Efrem, 17 anni)

Caro Efrem, la rivoluzione alla quale ti riferisci ci ha portati alla possibilità, grazie all'utilizzo di strumenti statistici e computer sofisticati, di analizzare in maniera veloce e automatica un'incredibile quantità di dati, estraendone valore con dei risultati fino ad ora inimmaginabili. Oggi la massa di dati generata è enorme e le aziende, come anche le persone, stanno diventando sempre più delle fabbriche di questi ultimi. Ognuno di noi ogni giorno lascia dietro di sé una densa traccia d'informazioni. Per capire di che cosa stiamo parlando, basta pensare alle carte di credito utilizzate per gli acquisti o alle foto che postiamo sui social. Queste miniere di dati, se analizzate con i metodi odierni, permettono di generare nuova conoscenza e sviluppare dei modelli probabilistici utili a creare delle previsioni su quello che potrebbe accadere in futuro. I Big Data offrono alle aziende che li sanno utilizzare un grande vantaggio competitivo e sono utili nei settori economici più disparati. Alcune aziende decidono le offerte migliori da proporre

ai loro clienti analizzando i comportamenti d'acquisto online. Nel campo medico questi metodi possono servire ad evidenziare gli effetti delle terapie a lungo termine, mentre a livello politico offrono un valido supporto per generare delle ipotesi sull'evoluzione delle elezioni.

Il lavoro del ricercatore di dati è indiscutibilmente complesso, intellettualmente intenso e non facile da inquadrare perché è una professione ancora giovane che manifesta molteplici sfaccettature provenienti da altre discipline scientifiche. Per lavorare in questo settore sono necessarie differenti competenze. In primo luogo bisogna possedere delle basi robuste in discipline quali l'informatica, la matematica, la fisica o le scienze computazionali; in secondo luogo sono richieste indubbie doti personali come l'intuizione, il ragionamento logico e tantissima creatività per edificare nuovi percorsi teorici e ideare soluzioni innovative. Attualmente, le università e i politecnici elvetici propongono molteplici percorsi di livello master o di formazione continua CAS, DAS e MAS che preparano all'esercizio di tale professione.

Link utili

Informazioni sui diversi indirizzi di studio offerti in Svizzera: www.orientamento.ch/studi

Informazioni sugli studi nel campo dell'informatica: www.orientamento.ch/informatica