

Rendere concreta l'informazione per elaborarla

Make information concrete for processing purposes

Violetta Lonati Mattia Monga

L'elaborazione automatica (cioè tramite una macchina) dell'informazione dipende dalla possibilità di rappresentare quest'ultima tramite configurazioni concrete della macchina che corrispondono a entità del mondo reale; l'elaborazione opera cambiando in modo coerente tali configurazioni. Nella maggior parte dei casi ciò comporta semplificazioni e perdite di informazione, cosicché una rappresentazione adatta a uno scopo non è lo è necessariamente per altri. L'elaborazione avviene tramite macchine generiche, capaci di compiere ogni possibile elaborazione: perciò la descrizione dell'elaborazione stessa deve essere rappresentata nella macchina, rendendola perciò suscettibile di ulteriori elaborazioni.

Parole chiave

informazione; rappresentazione; dati; codifica; macchine universali

Automatic (i.e., by a machine) processing of information depends on being able to represent it by concrete machine configurations that correspond real-world entities; processing operates by consistently changing these configurations. In most cases this involves simplifications and loss of information, so that a representation that is suitable for one purpose is not necessarily so for others. Processing is done by generic machines, capable of doing any possible processing: therefore, the processing itself must be represented in the machine, thus making it amenable to further processing.

Keywords

information; representation; data; coding; universal machines

 **Corresponding author:** mattia.monga@unimi.it

1. Introduzione

Agiamo, ormai da tempo, in contesti sociali, produttivi, affettivi in cui gli aspetti guidati o influenzati dall'elaborazione **automatica** delle informazioni sono in continua crescita. Non a caso, la chiamiamo sempre più spesso la "società dell'informazione", dimenticando forse che ciò che davvero distingue l'evo attuale non è tanto la grande disponibilità di informazioni, ma il fatto che esse siano suscettibili di manipolazione automatica, con una pervasività e ampiezza mai raggiunta prima. Secondo il filosofo L. Floridi¹ viviamo in un mondo in cui l'informazione non è semplicemente il riflesso di ciò che accade, ma ne determina la natura: il mondo dipende, a un livello senza precedenti, dalla sua **rappresentazione**. Sarebbe quindi assai opportuno che una riflessione su questi temi iniziasse fin dal primo ciclo d'istruzione, affinché nel secondo ciclo si possano approfondire alcuni degli aspetti più interessanti dell'impatto dell'informatizzazione della società e della scienza.

2. Rappresentare (concretamente) per elaborare

Affinché una elaborazione possa essere portata a termine **automaticamente**, cioè eseguita da un dispositivo ("macchina") di cui siamo in grado di predire con certezza sufficiente il comportamento, l'informazione da elaborare (dati di ingresso, o *input*) deve essere rappresentata da **entità concrete**. Durante l'elaborazione, tali entità vengono manipolate e **cam-biate** dalla macchina in maniera coerente a ciò che esse rappresentano; al termine dell'elaborazione lo stato di queste entità, interpretato sempre in base al sistema di rappresentazione usato, fornirà l'informazione attesa (dati in uscita, o *output*). Ciò può svolgersi tramite rappresentazioni analogiche; ad esempio una sveglia elabora informazione temporale rappresentata tramite angoli analoghi allo scorrere del tempo (operati da ingranaggi e visualizzati tramite lancette): ogni configurazione del dispositivo **rappresenta** uno specifico orario e viceversa; la sveglia scatta nel momento in cui il sistema entra nella configurazione corrispondente all'orario scelto, come se il sistema "riconoscesse" (è facile trovare appropriati termini antropomorfi quando si descrive un processo di elaborazione dell'informazione, che è del resto un'attività comune anche ai nostri processi cognitivi) la rappresentazione dell'orario fissato.

Risultano tuttavia particolarmente efficaci i sistemi in cui l'informazione è resa in maniera **simbolica**, tramite l'uso di alfabeti finiti di simboli: in questo è paradigmatico l'uso che facciamo delle cifre (*digit* in inglese, i simboli 0,1,2,3,4,5,6,7,9) per rappresentare il concetto astratto di numero, infatti questo tipo di rappresentazioni sono spesso dette **digitali**. Sorprendentemente, è possibile costruire automatismi **universali**, in grado cioè, in linea di principio, di eseguire —con lo stesso meccanismo!— tutte le elaborazioni effettivamente realizzabili. Questa è una conquista scaturita soprattutto dal grande sforzo intellettuale della cosiddetta "crisi dei fondamenti della matematica", a cavallo fra la fine del XIX e l'inizio del XX secolo, grazie alla quale l'informatica può trascurare quasi completamente il meccanismo e concentrarsi invece sul processo astratto reso possibile da esso. Rimangono quindi due punti di grande interesse teorico e pratico:

¹L. Floridi, *La quarta rivoluzione*, Raffaello Cortina Ed., 2017

1. Rappresentare l'informazione.
2. Descrivere l'elaborazione.

In questo articolo ci concentriamo soprattutto sul primo punto.

3. Informazioni e dati

Nel linguaggio comune, la parola ‘informazione’ può essere usata per riferirsi a qualcosa che prima non si sapeva, qualcosa che aumenta la nostra conoscenza o che riduce la nostra incertezza. Chiediamo informazioni sulla strada da prendere o su un fatto avvenuto; ci informiamo di ciò che accade nel mondo ascoltando il giornale radio, o consultando le notizie su un quotidiano o su un sito di notizie *online*. Tra le numerose definizioni alternative di informazione raccolte nel glossario del *National Institute of Standards and Technology* (NIST), una delle più ampie è la seguente²:

Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning.

L'informazione solitamente prende forma e si esprime tramite **dati**, elementi basilari di conoscenza, che possono essere osservati, memorizzati, elaborati. Di per sé, i dati sono privi di un significato intrinseco e necessitano quindi di essere interpretati nel contesto affinché forniscano informazione (che informazione ci dà il numero 42? Potrebbe essere l'età di una persona, il numero di pezzi contenuti in una scatola di cioccolatini, ecc.). Un dato può essere informativo per qualcuno ma non per altri, poiché l'informazione che il dato esprime dipende dal contesto e da chi lo riceve e interpreta, e in particolare dalle sue conoscenze pregresse.

Viceversa, nel decidere come convogliare un'informazione tramite dati, è sempre necessario **compiere delle scelte** e c'è quasi sempre una **perdita di informazione**. Il processo è intrinsecamente suscettibile di errore, o quanto meno di approssimazione. Per dare informazioni sull'età di una persona, ad esempio, si può scegliere di indicare quanti anni ha (42) oppure fornire la sua data di nascita (12 gennaio 1983); il primo dato in molti casi può essere sufficiente per convogliare l'informazione sull'età, ma non ad esempio per poter sempre individuare chi è più giovane tra due persone.

La scelta di quali dati usare per esprimere un'informazione dipende dunque dal contesto e dall'uso che se ne vuole fare. Immaginiamo di voler distinguere tra due figure: un rettangolo rosso e un triangolo equilatero giallo. Ci basterà ad esempio dire quanti lati ha la figura scelta (3 o 4) oppure dire di che colore è (rosso o giallo). In entrambi i casi sceglieremo di concentrarci su un aspetto (ad esempio il numero di lati) e di **astrarre** da altri aspetti (ad esempio il colore) che dunque considereremo secondari, se non irrilevanti. Il dato scelto sarà comunque in grado di convogliare l'informazione necessaria, a patto che chi lo riceva conosca già quali sono le due possibili figure alternative tra cui scegliere. Naturalmente invece questo dato da solo non basterebbe se lo scopo fosse quello di far disegnare la figura

²<https://csrc.nist.gov/glossary/term/information>

scelta ad una persona che non l'avesse già vista e che, pur avendo saputo che i lati sono 3, potrebbe ad esempio disegnare un triangolo non equilatero di altro colore o dimensione. Queste considerazioni potrebbero apparire ovvie, se non incontrassimo così frequentemente casi in cui questi aspetti sono stati chiaramente ignorati, con conseguenze spiacevoli. Vediamo un esempio (per la versione completa si rimanda a Monga³, 2024). Il 12-13 febbraio 2023, nelle elezioni per il Consiglio Regionale della Lombardia si è votato secondo la Legge Regionale n.17/2012 come modificata dalla Legge Regionale n.28/2017⁴. In particolare, la norma dispone che:

Nel caso di espressione di due preferenze, esse devono riguardare candidati di sesso diverso della medesima lista, pena l'annullamento della seconda preferenza.

Elettori ed elettrici possono scegliere i candidati consultando il “Manifesto elettorale” di cui si fornisce copia a tutti i seggi elettorali. La figura 7 mostra quello della circoscrizione provinciale di Milano. Il manifesto riporta per ogni candidato/a il cognome (in maiuscoletto) seguito dal nome e, in caratteri più piccoli, il luogo e la data di nascita. Non vi è invece nessun dato che esprima il sesso (si intende quello registrato all'anagrafe) dei candidati e delle candidate. Si potrebbe pensare che quest'informazione sia già contenuta nel nome e dunque derivabile da esso, ma come si può apprezzare leggendo i nomi nella figura, quest'affermazione è assai discutibile. Il problema è di principio, visto che esistono nomi usati sia per i maschi che per le femmine (Andrea, Gabriele, ecc.), e diventa ancor più evidente con nomi meno diffusi tradizionalmente nel territorio geografico di cui si parla (ad esempio, qual è il sesso di CHANDRASEKARA Padukkage Dhammika Nilmin Priyadarshani?). Di fatto, i manifesti non riportano in nessun modo affidabile un'informazione essenziale ai fini della validità del voto. Il sesso dei candidati —caratteristica essenziale per esprimere validamente due preferenze— è in generale conoscibile, ma **non è riportato nel manifesto elettorale**. Si noti che il problema non è legato al fatto di volere compiere l'elaborazione con un automatismo: anche elaboratori umani sono in difficoltà: come possono gli elettori e, soprattutto, gli scrutatori valutare la validità delle preferenze espresse?

4. Codifica dei dati

Una volta scelto con quali dati si vuole esprimere un'informazione, si deve anche scegliere in che modo **rappresentare** questo dato (in termini tecnici, si parla di **codifica** del dato). Ad esempio il sesso potrebbe essere rappresentato con una dicitura maschio/femmina, con un simbolo o un'icona (come sulle porte dei bagni pubblici), con una sola lettera M/F o —più astrattamente— con un codice binario (0/1). Oppure potrebbe essere rappresentato tramite il codice fiscale (assieme ad altri dati, come data e luogo di nascita). Naturalmente per **decodificare** il dato rappresentato è necessario conoscere il **sistema di codifica** usato (ad

³M. Monga, *Non dato, ma creato*, Mondo Digitale, in corso di pubblicazione.

⁴<https://normelombardia.consiglio.regione.lombardia.it/NormeLombardia/Accessibile/main.aspx?view=showdoc&iddoc=lr002017122800038>

esempio 0 = femmina, 1 = maschio o viceversa: la codifica è, per definizione, completamente arbitraria).

Un po' come per una lingua, in cui migliaia di parole possono essere scritte usando sequenze di lettere dello stesso alfabeto di qualche decina di simboli, è possibile codificare dati di tipo diverso (numeri, testo, immagini, suoni, ecc) come sequenze di simboli, garantendo la possibilità di decodificare senza ambiguità i dati rappresentati. Portando all'estremo questo approccio, e a patto di utilizzare sequenze di simboli più lunghe, si può arrivare a definire codifiche che usano un alfabeto di soli due simboli (tradizionalmente lo 0 e l'1): di fatto, i dati che vengono manipolati dai dispositivi digitali che usiamo quotidianamente sono codificati in questo modo tramite dispositivi capaci di assumere due diverse configurazioni (per questo la codifica binaria è così importante). Questo non significa però che gli informatici lavorino in binario! Nella loro attività quotidiana, ad esempio nell'attività di scrittura di programmi (che a loro volta forniscono la descrizione di metodi di elaborazione dell'informazione) si fa infatti riferimento a rappresentazioni e codifiche più facilmente comprensibili da chi scrive e legge i programmi stessi, affidando a ulteriori elaborazioni automatiche la traduzione nelle codifiche effettivamente necessarie al funzionamento del sistema informatico. Questo ultimo passaggio è in realtà cruciale: i programmi per l'elaborazione dei dati, sono essi stessi "dati", cioè la rappresentazione di un'informazione (relativa a come deve avvenire l'elaborazione).

5. Strumentalità delle rappresentazioni

La scelta di come rappresentare un dato (così come la scelta di quale dato usare per esprimere un'informazione) è anch'essa non scontata e produce effetti relativi alla facilità di leggere, interpretare ed elaborare il dato stesso: individuare il sesso di una persona dal suo codice fiscale non è immediato, e richiede la conoscenza di alcuni dettagli su come sono strutturati i codici fiscali; d'altra parte questi forniscono indubbi vantaggi nel contesto di un'elaborazione automatica (ad esempio disponendo di un carattere finale di controllo che permette di rilevare eventuali errori di trascrizione).

Nel caso di altri dati, come quelli numerici, è chiaro che la **rappresentazione decimale** (p.es. 12) risulterà comoda se vogliamo fare delle operazioni aritmetiche, almeno rispetto all'uso dei nomi dei numeri ('dodici') o alla **rappresentazione romana** (XII). Quest'ultima potrebbe però avere altri vantaggi (dopotutto è stata usata per secoli): forse, per esempio, è più adatta alla scrittura per incisione⁵. La **rappresentazione binaria** standard (12 sarebbe rappresentato come 1100) potrebbe essere particolarmente efficace se la nostra elaborazione è dedurre la parità del numero, dato che per tutti i numeri pari la cifra più a destra nella rappresentazione binaria è 0. In breve, una rappresentazione può facilitare una elaborazione e renderne difficili (o addirittura impossibili) altre.

Anche se in parte questo tipo di ragionamenti dovrebbe già far parte del panorama dei saperi esplorati a scuola (in particolare grazie alla riflessione fra modelli matematici della realtà e la loro utilità nella pratica), l'approccio informatico permette una operatività diretta inusuale, rendendo forse più evidente la necessità di una specifica competenza sul tema.

⁵C.A.R. Hoare, C.B. Jones. *The mathematics of programming. Essays in computing science*. Prentice-Hall, Inc., USA, 1989

La possibilità di effettuare elaborazioni automatiche forza ed esplicita la rappresentazione dell'informazione ed evidenza come astrazioni rappresentative sbagliate abbiano effetti molto concreti nell'efficacia delle elaborazioni.

6. L'elaborazione va rappresentata e può essere essa stessa elaborata

Per concludere, torniamo su un concetto introdotto *en passant* alla fine del paragrafo 4: l'elaborazione **deve** essa stessa avere una rappresentazione esplicita nella macchina. Ciò deriva direttamente dalla **universalità** del principio di funzionamento su cui si basano (per esempio, la cosiddetta "macchina di Turing", un meccanismo ideale proposto da Alan M. Turing (1912-1954) nel 1936⁶). Il meccanismo *hardware* può sostanzialmente adattarsi a compiere qualsivoglia elaborazione automatica. Ma allora, se, in potenza, la macchina può portare a termine **qualsiasi** elaborazione, bisogna avere un modo per indicare quella **specifica** che ci interessa. E il modo è, in definitiva, rappresentarla tramite un **programma informatico**; il programma diventa dunque parte dei "dati" su cui lavora il meccanismo universale. Questa sostanziale intercambiabilità fra dati ed elaborazione è chiaramente una grande opportunità ingegneristica, ma anche una delle ragioni per cui il nostro intuito va allenato per poter ragionare utilmente sui sistemi informatici. Può un programma progettato per suonare canzoni ordinare un bonifico dal nostro conto corrente? Il nostro intuito probabilmente ci suggerirebbe una risposta negativa: in fondo si tratta di una elaborazione che trasforma dati musicali in dati sonori. Ma le macchine informatiche sono spesso capaci di elaborare (magari al di là delle intenzioni del loro progettista) i dati in modo da cambiare completamente gli effetti dell'elaborazione stessa, e una musichetta creata ad arte potrebbe provocare un trasferimento di denaro. Il cambio è sorprendente solo per un osservatore che non l'ha previsto, poiché, in realtà, dal punto di vista informatico è semplicemente una attualizzazione alternativa della intrinseca potenzialità universale. Uno degli attacchi più usati dai criminali informatici, il cosiddetto "*stack smashing*" (o "*buffer overflow*") si basa in effetti proprio su questa idea⁷.

Acquisire una certa intima familiarità con la natura⁸ delle elaborazioni informatiche (da non confondere con la semplice esperienza delle loro applicazioni) è dunque essenziale per orientarsi nella società in cui viviamo e l'opportunità di svilupparla dovrebbe essere proposta a tutta la popolazione studentesca e non, come oggi accade, solo nei percorsi specializzati.

⁶A. M. Turing, *On computable numbers, with an application to the Entscheidungs problem*. Proceedings of the London Mathematical Society Series/2 (42) (1936): 230-42. Per una presentazione divulgativa della macchina di Turing si veda: M. Monga, D. Malchiodi, A. Morpurgo, M. Torelli *Turing: la nascita dell'intelligenza artificiale*, Ed. Corriere della Sera, 2017

⁷Aleph One, *Smashing The Stack For Fun And Profit*, Phrack Magazine 7/49 (1996). Per una presentazione divulgativa si veda: M. Monga *È dell'informatica il fin la meraviglia. Metafore per la sicurezza e sicurezza delle metafore*, Informatica Umanistica 4, 2011

⁸V. Lonati, C. Mirolo, M. Monga Di cosa parliamo quando parliamo di 'programmi', Mondo Digitale 20(97), 2022

ELEZIONE DEL CONSIGLIO REGIONALE E DEL PRESIDENTE DELLA REGIONE LOMBARDIA DI DOMENICA 12 E LUNEDÌ 13 FEBBRAIO 2023

CIRCOSCRIZIONE PROVINCIALE DI MILANO

CANDIDATI ALLA CARICA DI PRESIDENTE DELLA REGIONE E LISTE PROVINCIALI COLLEGATE PER L'ELEZIONE DI N. 25 CONSIGLIERI

1 AREA FONTANA					2 Performance MILANO				3 Area D'AMORE	4 Letizia Moratti PROGETTO EUROPEO della letizia MORATTI	
LESTE PROVINCIALE COLLEGATE					LESTE PROVINCIALE COLLEGATE				LESTE PROVINCIALE COLLEGATE	LESTE PROVINCIALE COLLEGATE	
LESTE 1	LESTE 2	LESTE 3	LESTE 4	LESTE 5	LESTE 6	LESTE 7	LESTE 8	LESTE 9	LESTE 10	LESTE 11	LESTE 12
LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12	LESTE 1 LESTE 2 LESTE 3 LESTE 4 LESTE 5 LESTE 6 LESTE 7 LESTE 8 LESTE 9 LESTE 10 LESTE 11 LESTE 12

Figura 1: Il manifesto elettorale della circoscrizione provinciale di Milano (Elezioni regionali 2023)

Violetta Lonati, Mattia Monga
Università degli Studi di Milano