



Questo testo sviluppa un aspetto particolare
delle osservazioni contenute nel testo originale
– e anche nei commenti all'edizione italiana –
di *How to Lie with Statistics* di Darrell Huff.

<http://htlws.it>

Le “statistiche” dei siti web

Giancarlo Livraghi – maggio 2007

Un caso particolare, rispetto al mondo generale delle statistiche, riguarda i siti web. Alcune osservazioni su questo argomento possono essere interessanti per l'uso di quelle specifiche analisi, ma anche come esempio del fatto che i dati hanno significati e valori diversi secondo il modo in cui sono interpretati – e questo riguarda ogni genere di statistiche o di analisi su basi numeriche.

Si tratta di quel sistema che permette di valutare l'affluenza di persone su un sito online e il loro utilizzo dei contenuti. Si basa su software che analizzano i *log*, cioè i dati di accesso, e ne deducono “statistiche” secondo diversi parametri. Gli indici possono essere molti e complessi. Qui parliamo di tre fra i più rilevanti: il numero di *user session* (cioè di persone che visitano il sito), di *pageview* (cioè di pagine “viste” – anche se non necessariamente “lette”) e di *hit* (cioè di accesso a singoli file).

Quest'ultimo dato è il più alto, ma anche il meno rilevante, perché un singolo documento (“pagina”) può contenere un numero molto variabile di “pezzi” (per esempio immagini) che per il computer e per la rete sono “atti di connessione” ma dal punto di vista della lettura sono impercettibili e non danno alcuna indicazione utile per valutare l’attività delle persone che si collegano.

Alcuni anni fa c’erano molti casi di errore, perché il dato di *hit* era interpretato per ciò che non è – e assurdamente paragonato ad altri indicatori (come, per esempio, lettori di un giornale o spettatori televisivi). Quegli errori portarono a valutazioni grottesche, talvolta anche a investimenti insensati. Ma oggi le valutazioni più diffuse sono meno ingenue e quel genere di disorientamento è abbastanza raro.

Può sembrare assurdo, ma è un caso vero. Otto anni fa, mentre infuriava la fallimentare batraciomachia dei “portali”, i proprietari di un grande gruppo editoriale italiano furono informati del dato di *hit* riguardante un loro sito.

Per qualche motivo (probabilmente un eccesso di cosmesi) il numero era estremamente alto. Senza minimamente approfondire il dato, pensarono che quella cifra si potesse paragonare al numero di lettori delle loro testate.

Ne dedussero con un calcolo frettoloso una stima di possibili entrate pubblicitarie. Spaventati dal timore di “perdere l’autobus” (come avevano fatto, vent’anni prima, con la televisione) approvarono in quattro e quattr’otto un piano di investimenti (comprese acquisizioni) che negli anni seguenti si tradusse in un enorme, quanto inutile, spreco di denaro.

Questo è tutt’altro che un caso isolato – e avventure altrettanto catastrofiche ci furono anche in altri paesi. Un’ennesima conferma di quanto possa essere pericoloso “mentire con le statistiche”.

Fermiamoci perciò su due parametri più semplicemente rilevanti – e (se bene usati) più significativi.

Il numero di *user session* (in un giorno, un mese o un anno) ci dice quante persone sono “venute a vedere” il sito (o uno dei suoi contenuti). Naturalmente si può trattare di tante persone diverse che “accedono” poche volte o di un minor numero di visitatori che ritornano più spesso. Ma come dato totale è una misura significativa del “traffico”.

Il numero di *pageview* ci dice quante “pagine”, cioè quanti singoli documenti (lunghi o brevi) presenti nel sito sono stati “visti” da quelle persone.

Va ricordato che il concetto di “pagina” in un sito web è molto diverso dalle pagine di un giornale, una rivista o un libro. Una “pagina” online può essere più lunga o più breve di una pagina o di un capitolo di un libro. Equivale concettualmente a un “articolo”, che in un giornale o in una rivista può occupare uno spazio relativamente piccolo oppure più pagine. Online, naturalmente, è possibile scegliere se offrire un testo lungo in un’unica “pagina” oppure separarlo in parti collegate con *link*. In generale è meglio non avere “pagine” troppo lunghe, ma la scelta della soluzione migliore può variare molto secondo la natura dell’argomento e la struttura del sito.

“Se ne deduce” abitualmente che un sito ha tanto più successo quanto più alto è il numero di visitatori – e che quei visitatori sono tanto più interessati quanto più alto è il numero di “pagine” viste. Ma la cosa non è così semplice.

In un interessante articolo pubblicato il 28 maggio 2007 *Web’s key management metric: task completion*¹ Gerry McGovern (un autore molto qualificato ed esperto su questi argomenti) ci fa notare come degli stessi dati si possano dare diverse interpretazioni.

Se una persona ritorna ripetutamente su un sito, osserva Gerry McGovern, può essere perché non riesce, con il primo tentativo, a svolgere il compito (*task*) che desidera. Se visita più pagine, può essere perché le prime che trova non corrispondono a ciò che sta cercando. Un servizio ben concepito dev’essere in grado di soddisfare le esigenze dei lettori nel modo più semplice possibile, con il minor numero di passaggi e nel tempo più breve.

Molti siti hanno sistemi mal concepiti, inefficienti, faticosi. Bizzarramente potrebbero considerarsi “premiati” da un maggior numero di visite, e di pagine viste, dovuto alle difficoltà che stanno creando. (La cosa, ovviamente, è grave quando si tratta di servizi pubblici, o relativi a una particolare organizzazione, perciò il lettore non può semplicemente abbandonare il sito ostico e cercare migliori esperienze altrove).

In senso del tutto opposto, un sito interessante ed efficace può avere molte visite, e molte pagine lette, perché offre contenuti validi e utili approfondimenti.

¹ <http://www.gerrymcgovern.com/nt/2007/nt-2007-05-28-task2.htm>

Insomma gli stessi dati possono avere, secondo il caso, significati diversi. Non basta leggere superficialmente le statistiche secondo schemi semplicistici e standardizzati. Occorre molta più attenzione nel capire le relazioni e valutare la qualità dei contenuti, e della loro organizzazione, dal punto di vista di chi legge.

Può essere opportuno ricordare qui il concetto di comunicazione in rete come è riassunto alla fine di *Le imprese e l'internet* (<http://gandalf.it/upa/>).

Quando le tecnologie sono al servizio dei contenuti, e le une e gli altri sono al servizio di chi legge o di chi cerca qualcosa in rete, l'esperienza dell'utente (lettore, cliente, interlocutore) è fluida, efficiente, gradevole; porta rapidamente e senza difficoltà a ciò che quella persona cerca o desidera. Questa è la chiave fondamentale di ogni efficace comunicazione online.

Si tratta di valori e qualità che non è possibile misurare solo con indici numerici. Le valutazioni statistiche possono essere utili, se bene impostate e interpretate. Ma non possono mai sostituire le esperienze dirette dell'interattività e la comprensione dei valori umani.

Non solo in questo caso, ma in ogni genere di dati e statistiche, è sempre necessario capire bene i valori qualitativi – e tener conto delle conoscenze concettuali e culturali, di sensibilità e di metodo, che permettono di impostare meglio l'analisi e di interpretare più utilmente i risultati.