

## L'Italia dell'Unità

# Un «occhio elettronico» ha memorizzato più di 548mila pagine

**Il progetto** 87 anni di edizioni del giornale sono state indicizzate attraverso il «riconoscimento ottico dei caratteri» anche delle versioni in microfilm. Il lavoro curato da Tiscali ha prodotto una mole di dati pari a 225 gigabytes

**GIUSEPPE RIZZO**

Il primo numero dell'Unità è uscito il 12 febbraio 1924, costava 20 centesimi e contava quattro pagine fitte di caratteri duri e neri e minuti – fotografie: zero. Il giornale che state sfogliando oggi indossa inevitabilmente un altro vestito e lo trovate, oltretutto in edicola, online, su iPad e iPhone. Ci sono, a dividere questi due quotidiani, 87 anni di storia italiana – editoriale, culturale e politica. L'Unità di oggi è però riuscita a costruire un ponte grazie al quale attraversare queste sette decadi con pochi semplici click. Un archivio digitale elaborato dal gruppo Tiscali, e consultabile gratuitamente su [Unita.it](http://Unita.it), in cui sono state raccolte le edizioni che vanno dal 1924 al 2008 – dal 2008 le edizioni sono già disponibili in formato digitale nella sezione edicola di [Unita.it](http://Unita.it). Un'enorme mole di dati attraverso cui è semplicissimo navigare grazie a un sistema di ricerca che consente di raggiungere le singole edizioni ma anche le parole ricercate in ognuna di esse.

Per capire di cosa stiamo parlando, probabilmente è utile fare qualche numero. Le pagine elaborate ed indicizzate sono state complessivamente 548 mila 371, per una quantità di dati iniziali pari a 225 Gigabytes. Volendole pensare fisicamente, queste pagine, dovremmo immaginare una palazzina di qualche piano interamente costruita con l'Unità – fondazioni comprese. Per un risultato finale che aggiunge al valore storico-documentale dell'operazione, quello di ricerca e avanguardia tecnologica.

Le tecniche utilizzate per costruire l'archivio sono state diverse, e vale la pena tratteggiarle per

cogliere il senso dell'impresa. Le 289 mila 423 pagine che separano il primo numero dell'Unità fino all'ultimo del 1996 sono state elaborate attraverso un processo di «riconoscimento ottico dei caratteri». Ovvero grazie a un software che consente l'interpretazione dei carat-

teri presenti all'interno di un'immagine (nella fattispecie: nei microfilm in cui erano conservate le edizioni dell'Unità) e ne permette la loro conversione in testo. Il risultato è una quantità di materiale pari a 1.8 terabytes e a 2.8 milioni di file. Per aggiungere vertigine a vertigine, ba-

sta pensare che durante questa operazione sono stati processati oltre 6 miliardi di caratteri. Per gli anni che vanno invece dal 1996 al 2008 sono stati generati ulteriori 2.7 milioni di file (420GB).

Per consentire la massima rintracciabilità in questa giungla di bit «sono state riconosciute in modo specifico l'edizione del giornale (Nazionale, Firenze, Bologna, Roma, Torino, Meridionale) l'anno ed il numero di uscita, la data e il numero di pagina – spiega Domenico Dato, responsabile Tiscali del progetto –, inoltre sono stati utilizzati degli algoritmi ad hoc per il riconoscimento dei titoli all'interno delle pagine, in modo da migliorare il ranking». Ed ecco, qualunque cosa sia il ranking, qualunque significato voglia dargli una persona all'oscuro dei meccanismi dell'informatica, bisognerà pure ringraziarlo, se è vero che grazie ad esso riusciamo a leggere alcune delle più grandi firme del giornalismo, della politica e della cultura italiana in un unico grande giornale che si srotola sotto i nostri occhi grazie a una semplice connessione internet.



I funerali di Enrico Berlinguer con la storica prima pagina «Addio»