

Electronic Document Management System

Il settore delle memorie ottiche è pieno di sfaccettature e di applicazioni, dalle più semplici e prevedibili, quali i CD-ROM con i programmi shareware e public domain, a quelle più inaspettate che sconfinano nell'intelligenza artificiale, sino alla rivoluzione nel campo dell'ufficio e dell'archivio

Il motivo che ha portato alla nascita dei sistemi a registrazione ottica è essenzialmente quello della necessità di una grande quantità di dati disponibili on-line. Una soluzione di tipo magnetico per 1 o 2 Gigabyte ha decisamente costi astronomici, mentre l'equivalente ottico no. Questo comporta perciò una serie di situazioni derivate, tali da fare nascere una nuova disciplina: l'EDMS ossia Electronic Document Management System.

Per comprendere meglio cosa sia l'EDMS è bene fare un piccolo passo indietro, partendo proprio dalla motivazione della nascita del computer; lo scopo primario era proprio quello di «computare», quindi calcolare grosse quantità di numeri che avrebbero richiesto altrimenti anni-uomo. A questa applicazione principe si è parallelamente aggiunta quella dell'archiviazione dati e del successivo data-processing. Archi-

viare è dunque una delle applicazioni più importanti del nostro settore, perciò la logica evoluzione della specie prevede l'uso di memorie sempre migliori quali supporti per i dati da archiviare.

Oggi le memorie ottiche possono vantare determinate caratteristiche (costi, capacità, affidabilità) che non si riescono a riscontrare in altre memorie di massa, pertanto hanno permesso lo sviluppo di un nuovo concetto di archiviazione, quello «paperless» cioè senza carta.

Il problema

Archiviare tramite elaboratore significa essenzialmente acquisire dati tramite tastiera e altro mezzo di input e di inserirli in un DBMS, il quale tramite appositi programmi ci consente di ripescare quelli che soddisfano determinate caratteristiche, impostate al momento della ricerca. Ad esempio possiamo ricercare le fatture emesse dal 10/10/78 al 10/12/81 il cui numero progressivo inizi per 3 e che il destinatario abbia la sede a Roma o provincia; questo è il lavoro quotidiano di migliaia di elaboratori, i quali devono poter registrare i dati immessi in apposite memorie di massa, per poi poterli gestire in consultazione.

L'apparizione dei CD-ROM ha consentito di liberare gli hard disk di una notevolissima quantità dati: tutto l'archivio storico non più soggetto a variazioni si può tranquillamente pressare nel media a sola lettura. In tal modo si ha sempre la possibilità di consultare i dati on-line, ma si liberano i dischi magnetici da 550 MB di dati per ogni CD-ROM stampato...

L'avvento dei dischi WORM ha invece portato alla nascita dei primi veri e propri sistemi EDMS, i quali non si limitano all'archiviazione dei dati relativi ai documenti (nell'esempio precedente le fatture), ma anche e soprattutto dell'immagine. Proprio questo è il punto chiave delle applicazioni EDMS: la memorizzazione, perciò la nascita, di archivi di dati ed immagini ad essi correlate.

Questo nuovo, rivoluzionario, mezzo di gestione degli archivi permette non solo di ricercare le fatture necessarie, ma anche di vederle a video e di stam-



parle se desiderato, evitando perciò di conservare la carta.

EDMS

Il nuovo concetto che si è quindi generato è quello dell'archiviazione del documento, non più dei soli dati relativi. Dal punto di vista dell'utente finale questa evoluzione è una vera e propria manna dal cielo che, con poco lavoro in più in fase di acquisizione, consente un risparmio di spazio e di tempo spaventosi; infatti l'operatore di un sistema EDMS non dovrà solo battere i dati sulla tastiera, ma anche passare il documento originale attraverso uno scanner. Esiste quindi un passaggio in più rispetto all'iter tradizionale, ma che aggiunge in archivio una fotocopia elettronica del documento originale.

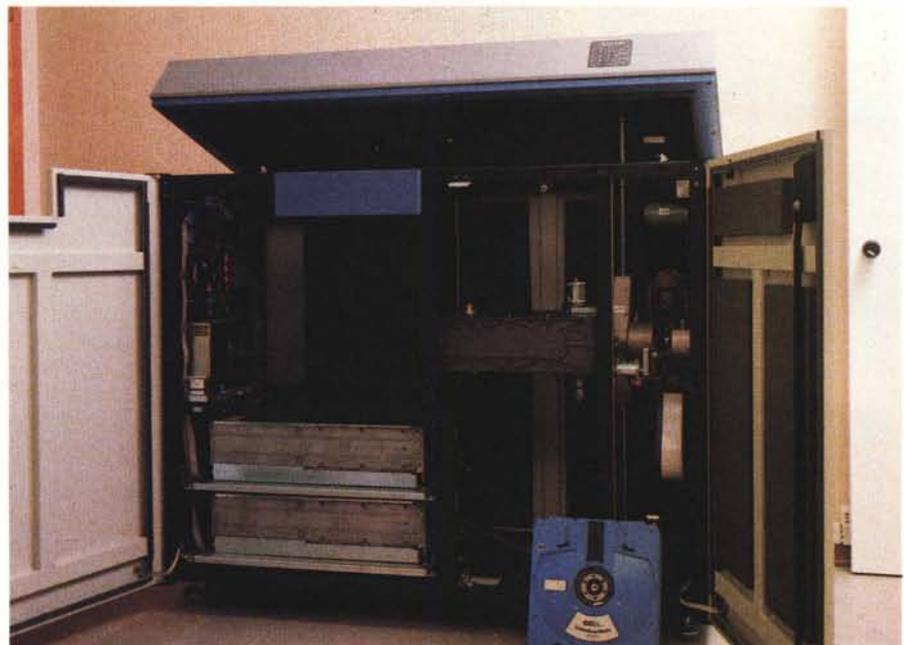
L'architettura di un sistema di questo tipo è naturalmente differente da quella tradizionale, in quanto il computer non userà solo monitor, tastiera e hard disk, ma anche delle altre, nuove, periferiche da gestire appositamente.

Il programmatore di un sistema EDMS si trova ad affrontare molti problemi in più, in quanto oltre al classico DBMS che resta la base del sistema, è necessario gestire tutte le periferiche ottiche, molto spesso lavorando in rete.

Questo fa quindi intuire quanto tali sistemi siano in piena fase di sviluppo ed espansione, ma che necessitano di un lavoro notevole sotto il guscio dell'interfaccia utente e dell'applicativo user friendly. La gestione elettronica dei documenti consente di automatizzare e di integrare moltissimi dei lavori d'ufficio, se sapientemente gestita: il primo settore ad essere rivoluzionato è quello dell'archivio standard: un sistema EDMS lo rimpiazza definitivamente in uno spazio enormemente più piccolo e con una efficienza tipica da sistema informatizzato. La gestione dei fax può essere collegata al sistema, automatizzando quindi anche quest'aspetto della vita dell'ufficio, ma le meraviglie del mondo EDMS consentono una totale e perfetta integrazione col settore dell'office automation, per una totale e completa automatizzazione delle varie funzioni.



Ecco il sistema EDMS «Laserview», basato su PC MS-DOS connesso ad un juke box.



Questo è l'interno di un juke box OSI a due drive: ciascun disco è in grado di contenere 2 Gigabyte di dati; in un juke box come questo c'è spazio per 17 dischi.

Tecnologie e macchine

La base dell'EDMS è quindi un Data Base Management System il quale non deve gestire solo i dati ma, soprattutto, anche le immagini dei documenti. Già da ora appare un dualismo che caratterizza questo ambiente: i dati e le immagini, due serie diverse di informazioni che seguono iter di lavorazione differenti. I dati sono la parte più consueta del mondo EDMS, vengono acquisiti e gestiti nella maniera più tradizionale: database e dischi magnetici, essenzialmente.

Le immagini sono invece una serie di dati binari un po' diversi dal solito: non sono programmi, senza dati servono a molto poco, occupano un sacco di spazio sulla memoria di massa e vanno gestite in maniera differente. Innanzi tutto dobbiamo fare un piccola parentesi riguardo la risoluzione delle immagini: la risoluzione è quel parametro che indica la quantità di micro dettagli, quindi informazioni, registrate, chiusa parentesi.

È chiaro che più alta è la risoluzione, più alto è il dettaglio, ma maggiore è la quantità di memoria necessaria alla registrazione.

Un ottimo livello di dettaglio per le immagini che debbono riprodurre documenti è il classico delle stampanti laser: 300x300 dot per inch (punti per pollice); il piccolo problema che un'immagi-

ne di qualità simile occupa circa 1 megabyte. Con due semplici conti si verifica che su una hard disk da 40 mega ci stanno a malapena 35 immagini, considerando 5 mega di dati utente: un po' pochine...

Ecco quindi focalizzato il primo problema: è necessario avere una memoria di massa molto capiente, in grado di ospitare le mie immagini, quindi che mi consenta di scriverle. La risposta in questo caso è data proprio dalla tecnologia di memorizzazione ottica: WORM. La scelta sul mercato mondiale è vasta, ma possiamo dire che i leader del mercato italiano sono Optotech (400 mega, 5" 1/4), Maxtor (800 mega 5" 1/4) e L.M.S. (2 giga, 12"), che corrispondono rispettivamente a 400, 800 e 2000 immagini: ancora poche.

È quindi necessario aggiungere una nuova prospettiva al problema: come ridurre l'ingombro delle immagini? La risposta ci viene da certi standard industriali usati per la costruzione delle macchine fax: metodi di compressione dati Huffman mono e bi-dimensionale; cosa significhi è presto detto: i dati componenti l'immagine vengono compressi secondo determinati algoritmi, riducendo l'ingombro in termini di memoria sino a 35 volte, a seconda del tipo di immagine. Questi due sistemi sono documentati dalle norme CCITT sugli standard di compressione fax III e IV. A questo punto si inizia ad avere su

WORM 14.000, 28.000 oppure 70.000 immagini: una bella quantità. La compressione è perciò un passaggio obbligato al momento della memorizzazione delle immagini, il punto che differenzia alcuni sistemi a basso costo da altri di costo maggiore è il modo in cui viene fatta la compressione. Il tipico metodo di compressione dei sistemi low cost è via software, quindi lento ed utilizza la CPU del computer per eseguire tale compressione. I sistemi EDMS più costosi incorporano invece una scheda che esegue via hardware, con un processore indipendente, la compressione dell'immagine in tempi molto più veloci, lasciando libera la CPU di eseguire qualcos'altro, l'acquisizione dati, per esempio.

Una volta ottenuta l'immagine compressa ed i dati relativi al documento, i due tipi di informazione prendono strade diverse: i dati vanno a finire nel DBMS sull'hard disk, mentre l'immagine sul WORM. Al momento di scrittura sul disco ottico, il software di gestione ritorna un codice identificativo dell'immagine, una sorta di nome file univoco, che ovviamente verrà memorizzato fra i dati nel DBMS, in modo da creare un concatenamento fra quei dati sul disco magnetico e quell'immagine sul WORM.

Naturalmente è necessario usare uno scanner per acquisire un'immagine, scanner in grado di far aumentare notevolmente il costo di un sistema: un conto è il classico HP in grado di leggere solo pagine in formato A4 in circa 20.5 secondi a 300 dpi (circa 5 milioni) un altro è il Ricoh IS-400 che legge una pagina A3 a 300 dpi in 3.5 secondi, con tanto di caricatore automatico in grado di tenere 35 fogli da leggere in sequenza (circa 30 milioni). La gestione di questa periferica è molto importante, perché bisogna poter settare da software i vari parametri relativi alla luminosità, la gestione dei grigi, la risoluzione ed il contrasto. Tutti questi controlli sono fondamentali per poter ottenere l'immagine con la migliore definizione possibile, il più vicino all'originale.

Per il controllo della qualità dell'immagine e per la successiva consultazione è importante avere un monitor che consenta la massima risolvibilità. È chiaro che lavorare col classico monitor Hercules, Ega o Vga non ha senso, poiché si avrebbe un'immagine molto più piccola del reale ed a bassa risoluzione. Per le applicazioni EDMS sono necessari i monitor A4 ad altissima risoluzione come minimo, essendo gli unici in grado di visualizzare a pieno formato reale i documenti acquisiti.

Per poter vedere le immagini già ar-



La Hewlett Packard è stata la prima azienda a fornire la sua documentazione tecnica su CD-ROM: il prototipo di un sistema EDMS.



Per convertire informazioni dal formato tradizionale a quello adatto al CD-ROM è necessario un sistema come il CD-Publisher della Meridian.

chivate è però necessario operare una decompressione per riportarle allo stato originale; anche in questo caso, come per la compressione, si può operare via software con tempi di attesa di circa 5-6 secondi o con schede hardware in grado di visualizzare la pagina in meno di mezzo secondo.

L'ultimo anello fondamentale della catena è la stampante: laser, assolutamente.

Non si può certo scivolare su una simile buccia di banana mettendo una stampante ad aghi: si perderebbe tutta la qualità del sistema e dell'acquisizione tramite scanner; una stampante laser è l'unica soluzione in grado di dare 300 dpi ad alta velocità, magari in copia multipla per poter dare ad ogni direttore una copia del documento «pseudo originale» richiesto, senza neanche fare una fotocopia... A questo punto si può già tracciare un esempio di rete EDMS di-

partimentale: un server dati di rete, un server immagini di rete, collegato al/ai disco/hi ottico/i, qualche stazione di consultazione documenti, quindi macchine con solo il video ad alta risoluzione, un paio di stazioni di acquisizione, perciò con scanner e video. Le stampanti possono essere locali o installate sui server, in modo di poter dare a tutti la possibilità di utilizzarle. Una simile soluzione consente di avere in circa 6 metri quadrati (la superficie occupata dal server immagini) una capacità di svariate migliaia (in certi casi milioni) di documenti, pronti ad essere visualizzati in 5-6 secondi da chiunque ne faccia richiesta...

Ultima frontiera

Non è finita qui: il mercato EDMS è considerato «very profitable» negli States, dove è stato valutato un giro di

centinaia di milioni di dollari l'anno con un incremento costante del 65% l'anno per 10 anni.

Alla luce di simili potenzialità di mercato, sono stati fatti discreti sforzi nella ricerca di integrazione di sistemi EDMS con altri sistemi di computer già esistenti in ambito aziendale (leggi IBM, Digital, Siemens) in modo di connettere i due mondi.

Questa ricerca è sfociata nella creazione di sistemi tipo COM (Computer Output to Microfilm) che però non portano lo spool di stampa su microfilm ma, indovinate un po', su disco ottico sotto forma di immagine, oppure l'immagine viene generata solo sul video al momento della richiesta di visionare quella stampa.

Per chi non si accontenta di avere solo 70.000 immagini su un disco, esistono appositi juke-box in grado di arrivare sino a 95 dischi ognuno (quindi 6.650.000 immagini di linea) con tempi di cambio disco di 10 secondi e di accesso all'immagine richiesta di 5 per un totale di 15 secondi circa, il tutto in metri 3x1.5x2.5: quanto spazio occorrerebbe per la stessa quantità di fogli di carta?

Per i pigri invece ci sono i sistemi OCR (Optical Character Recognition) ed ICR (Intelligent Character Recognition), il cui massimo rappresentante è il Palantir ES 9000, della ditta californiana Palantir.

È una belva in grado di ricevere le immagini tramite interfaccia seriale, o SCSI o Ethernet (tutte montate di serie) e di riconoscere i caratteri (quindi leggerli) alla velocità di 250 al secondo, indipendentemente dal font o dalla dimensione.

Integrato in una rete può ricevere le immagini dal server immagini, leggerne i dati scritti restituendoli alla stazione richiedente, la quale con un semplice programmino può aggiornare automaticamente il DBMS, senza neanche battere un dato...

Fine

Questa puntata è stata una generale panoramica di una «profitable» applicazione della seconda generazione nel campo delle memorie ottiche, con indirizzo WORM. In realtà i documenti storici possono essere portati su CD-ROM senza problemi, integrando così il sistema EDMS anche con questo tipo di memorie.

Se l'assaggio vi ha stuzzicato e vi interessa sapere cosa si muove in Italia in questo settore delle memorie ottiche, aspetto commenti e richieste nella conferenza CD-ROM su MC-Link. 



Le nuove tecnologie ottiche Erasable avranno presto spazio in questo settore.