

File in formato testuale

Come usarli in uno Spreadsheet

Nello scorso numero abbiamo parlato dei Linguaggi di Interrogazione presenti nei vari prodotti di Informatica Individuale, abbiamo parlato di prodotti Spreadsheet e di prodotti DBMS, mentre come linguaggi abbiamo parlato di SQL, QbE, QbB, e loro varianti. In questo articolo approfondiamo un tema appena accennato in quello precedente

di Francesco Petroni

Parliamo di come si possa, con uno spreadsheet, manipolare un file di dati, disponibile in formato testuale a lunghezza fissa (il termine tecnico è formato Fixed Length).

Riteniamo, ma lo dicono pure le statistiche, che lo spreadsheet sia il prodotto principale dell'Informatica Individuale (a nostro parere le sue regole fondamentali dovrebbero ormai essere insegnate nelle scuole secondarie). Riteniamo conseguentemente che un normale utilizzatore di un PC debba conoscere perfettamente tali regole, debba padroneggiare i comandi dello spreadsheet e debba, possibilmente, saper sfruttare anche i comandi d'uso meno frequente quando gliene capiti l'occasione.

Tra i comandi meno frequenti ci sono proprio quelli che permettono di «caricare» file dati disponibili in formati che

non siano quello proprio del prodotto. Il problema è quello di adattare alle «necessità» del foglio elettronico, che ha una sua organizzazione per righe e per colonne, i dati presenti nel file letto.

Il problema posto nei suoi termini più generali è quello di convertire un file dati disponibile in qualsiasi formato in un formato riconosciuto dallo spreadsheet che si sta usando.

In questo articolo parleremo solo di un tipo particolare di formato, molto diffuso perché è il più neutro e perché può essere realizzato con tutti i prodotti, il formato Fixed Length.

In particolare gli argomenti trattati saranno:

- come importare e come adattare alle celle del foglio un file FIXED LENGTH, in modo che i dati vengano ripartiti nelle celle;

- come leggere un file testuale, che però rimanga esterno al foglio, allo scopo di importarne solo alcune righe e alcuni campi, e allo scopo di manipolare i campi in fase di caricamento. Svolgeremo questo secondo argomento, più complesso, presentando delle Macro;

- varie. Ovvero altri sistemi, altri prodotti, altre considerazioni.

I prodotti che useremo saranno:

Lotus 123, in versione 3.4.

Microsoft Excel in versione 4.0

Borland Quattro Pro, nella Special Edition riservata ai lettori di MC.

Il problema della importazione e cosa è un File Fixed Length

Un file in formato spreadsheet, sia esso 123, Quattro, Excel, contiene oltre ai dati anche una serie di informazioni, ad esempio in quale cella risiede quel dato, quali eventuali formati numerici o estetici sono stati imposti a quella cella, ecc.

Importantissima è l'informazione «formula». Ovvero in un file formato spreadsheet sono memorizzate anche formule riferite a celle, del tipo +A1/B1. Un file dati invece non contiene altro che dati, non contiene formule, non contiene formati estetici.

Se si tratta di un formato dati «nobilito» riferito cioè ad un prodotto particolare di categoria DBMS, come dBase III o Paradox, viene conservata nel suo interno anche la «struttura» ovvero l'organizzazione del file stesso in campi. La struttura comprende nome, tipo e lunghezza del campo.

Ormai tutti i prodotti tipo Spreadsheet leggono e scrivono tali formati di file comprendenti dati. È chiaro che in caso

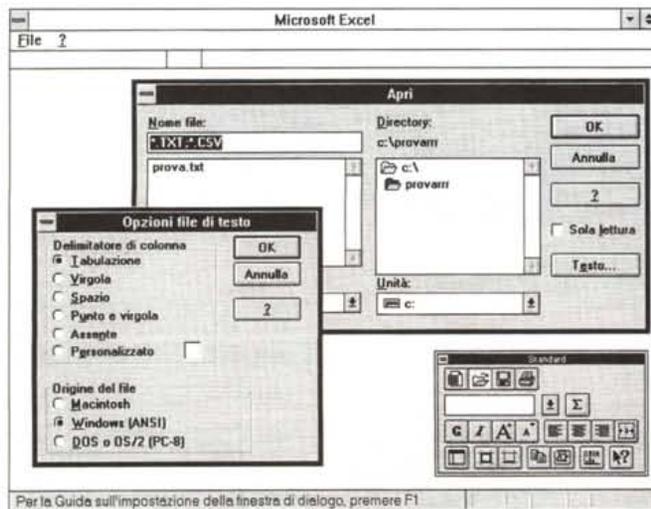


Figura 1 - Microsoft Excel 4.0 - Opzioni lettura di un File Testuale. Ormai tutti i prodotti di Informatica Individuale, destinati cioè al grosso pubblico, permettono di leggere file in formato testuale. In un file testuale ci sono solo dati. Non ci sono formule, non ci sono attributi estetici, non c'è indicazione di come vadano organizzati i dati una volta letti. È l'utente che deve conoscere, a parte, la cosiddetta struttura del record, ovvero la sua organizzazione. Utilissima è la presenza del file testuale dei separatori che fanno capire quando finisce un campo e ne inizia un altro.

mero in formato testuale e decimale. Ma questo è un problema che può capitare al programmatore che userà per la conversione il linguaggio e non lo spreadsheet.

I codici dell'Analisi

Finora abbiamo parlato di problemi generali comuni a tutti i prodotti. Nel passare alla specifica funzione di Analisi va detto che i prodotti di cui ci occupiamo si comportano in maniera leggermente differente.

La prima differenza sta nel modo di leggere il file testuale.

Lotus 123 e Quattro Pro leggono il file che va a finire sul foglio già attivo, in una zona scelta semplicemente definendo la cella a partire dalla quale posizionare i dati.

I comandi rispettivi sono Archivi Unisci e Varie Importa.

In Excel il file viene letto direttamente con il comando File Apri e quindi equivale all'apertura di un file Excel e tutti gli effetti. Se si vuole che tale file diventi Excel (ad esempio che passi da una designazione TXT a una XLS) occorre eseguire un File Salva Come e scegliere il tipo file Normale.

Per dare una parvenza di incolonnamento ai dati, letti ma ancora non ripartiti, anche nelle videate grafiche, la cosa migliore è quella di utilizzare un carattere a lunghezza fissa (il più noto è il Courier) che dà la stessa dimensione ai caratteri, ad esempio la M, la I e anche il Blank.

Una volta a bordo i dati, che occupano comunque una colonna di una sola cella, le funzionalità per ripartirli sono analoghe nei tre prodotti.

In 123 e Quattro occorre creare una Riga Campione che viene usata dalla funzione Analisi per ripartire i dati. In Excel invece la riga campione è la prima riga dati nella quale occorre inserire delle parentesi quadre che indicano inizio e fine del campo.

Prendendo i dati di prima avremo per 123 e Quattro:

```
E>>>>>E>>>>V>V>V>V>>>>V>
BIANCHIMARCO31125644400005
ROSSI UGO 111278 4560010
MARTINI 21116590000015
LA ROSAGUIDO010562 5000012
Mentre in Excel dovremo comporre la riga campione in questa maniera:
[BIANCHI][MARCO]
[31][12][56][444000][05]
```

I tre prodotti propongono delle righe campione basate su eventuali spazi vuoti trovati nella riga di partenza. Purtroppo è raro il caso in cui i vuoti di tale riga, che tralaltro potrebbe non esistere per niente, corrispondano esattamente a quelli

Figura 6 - Pioneer Software Q+E - Alle prese con un file Fixed Length. A beneficio di chi si ricorda quanto letto nell'articolo, presentato sul numero precedente di MC, riguardante i linguaggi di interrogazione, vediamo come il Q+E di Excel 4.0, permetta di definire, nel suo grosso comando SQL-Like, struttura e nome del campo del file Testuale letto.

CODICE	NOME	COGNOME	CITTA	FLAG	ANNO	MESE	GIORNO	QUAL	ALTRO		
1	1022	LUIGI	ROSSI	ROMA	S	1970	03	04	H	6538000	0.1
2	1027	COSSIMO	VERDI	MILANO	S	1956	11	24	B	1835000	0.1
3	1029	MARCO	BIANCHI	TORINO	N	1973	04	21	D	3999000	0.1
4	1034	MARIO	FUCSIA	FIRENZE	N	1964	01	10	H	6394000	0.1
5	1038	ALESSANDRO	GIALLI	NAPOLI	S	1971	09	21	C	1223000	0.1
6	1044	MARIANO	MARRONI	ROMA	S	1956	12	27	G	6059000	0.1
7	1047	LUDDOVICO	ROSA	ROMA	N	1953	06	03	B	3062000	0.1

Linea	Comando
1	ISTRUZIONI
2	ESTRAZIONE
3	Nome della Macro
4	Azzeraamento Variabili
5	Azzeraamento Variabili
6	Apertura del File Testuale
7	Legge una Riga del File
8	Scomposizione della Riga letta
9	Scomposizione della Riga letta
10	Scomposizione della Riga letta
11	Scomposizione della Riga letta
12	Scomposizione della Riga letta
13	Manipolazione della Riga
14	Scomposizione della Riga letta
15	Manipolazione della Riga
16	Scomposizione della Riga letta
17	Scomposizione della Riga letta
18	Manipolazione della Riga
19	Test per la Selezione
20	Riversamento del Dato sul Foglio
21	Riversamento del Dato sul Foglio
22	Riversamento del Dato sul Foglio
23	Riversamento del Dato sul Foglio
24	Riversamento del Dato sul Foglio
25	Riversamento del Dato sul Foglio
26	Riversamento del Dato sul Foglio
27	Riversamento del Dato sul Foglio
28	Riversamento del Dato sul Foglio
29	Manipolazione durante il Riversamento
30	Calcolo Dinamico della Media
31	Manipolazione durante il Riversamento
32	Incremento del Contatore
33	Riga successiva
34	Fine del Test
35	Torna alla Lettura Riga
36	Procedura di Chiusura
37	Procedura di Chiusura

Figura 7 - Microsoft Excel 4 - Una Macro che legge un file esterno. Una variazione sul tema Lettura di File Esterno. In questo caso la nostra Macro (in Excel le Macro occupano una speciale tipo di foglio) legge riga per riga il file esterno e lo scompone nei suoi vari campi, alcuni di questi sono ottenuti con dei calcoli. Il campo con la CITTA' viene sottoposto ad un test, in modo che vengano utilizzati (vengono riportati su un altro foglio) i soli dati relativi alla città di ROMA.

delle righe successive.

Come detto la Data è sempre un problema in quanto si legge come numero che poi si deve convertire. Si può leggere come un numero o una stringa unica o come tre numeri, giorno, mese e anno, e poi ricomporre con la funzione che, sulla base di tali tre valori, produce il numero seriale.

Sarebbe opportuno, prima di analizzare un grosso file testuale, fare delle prove, con un sottoinsieme di poche righe significative, nelle quali testare le situazioni più critiche.

In figura 3 vediamo una serie di test eseguiti con Quattro Pro SE, nei quali abbiamo provato i codici D che permettono di intercettare una Data, solo nel caso questa sia scritta in uno dei formati canonici, e che la traduce in un numero seriale. È evidente che poi nella colonna con la Data occorre affrettarsi a dare un for-

mato numero Data, che fa assumere al numero seriale il suo aspetto di data.

Nella stessa figura possiamo controllare quale siano i quattro passi logici dell'operazione.

Dapprima occorre posizionarsi sulla prima cella della colonna che contiene i dati. Partendo da questa cella si lancia il comando di Creazione della Riga Campione. In pratica il prodotto inserisce una riga nel foglio immediatamente sopra la cella da cui si è partiti ed in questa cella inserisce la riga di con i codici. Se questa non dovesse andare bene la si edita, con il comando Modifica che equivale a posizionarsi sulla cella incriminata e premere F2 (Edit).

Quando la riga campione è a posto si definisce la Zona con i Dati, che deve comprendere anche la riga campione e che deve essere larga una sola colonna. Si definisce poi la zona di Output, che se

Figura 10 - Borland Quattro Pro SE - Il manuale Utente. Lo abbiamo affermato più volte: la missione di MC è quella di propagandare la cultura dell'Informatica Individuale. L'iniziativa voluta da Technimedia e Borland del lancio di una versione speciale per i lettori di MC del Quattro Pro va in questa direzione. Si tratta di una raccolta di cinque fascicoli (ora acquistabili in una unica soluzione) che contiene il prodotto Quattro Pro SE, il materiale didattico, e cioè file e fascicoli con le Lezioni, il manuale ufficiale. Seguendo il corso si raggiunge un'ottima conoscenza del prodotto, che come noto è uno spreadsheet evoluto ed allineato, in termini di funzionalità a disposizione, con i maggiori concorrenti.



avesse ancora fatto di affrettarsi a procurarsi tale pacchetto, che, a un costo minimo, permette di avere un Prodotto Ufficiale (fatto che da diritto ad ulteriori vantaggi) e un corso completo che tratta in teoria e in pratica tutte le sue funzionalità.

Il nostro scopo, e anche l'iniziativa Quattro Pro SE va in questa direzione, è quello di propagandare la cultura dell'Informatica Individuale. Lo abbiamo detto più volte, per noi si tratta di una cultura di base, che andrebbe diffusa nelle scuole medie o medie superiori (i programmi ministeriali ancora non lo prevedono), e che va diffusa non solo perché la conoscenza di alcune tecniche dell'Informatica Individuale comunque serve nelle attività lavorative o nelle attività domestiche, ma anche perché l'esercizio di tali tecniche risulta formativo, nel senso che aiuta a mettere a fuoco le nostre innate capacità di logica.

Tornando al Quattro Pro SE va detto che si tratta di un prodotto spreadsheet molto evoluto e che contiene tutte le funzionalità presenti in tale categoria di prodotti.

Esistono prodotti successivi un po' più evoluti, che hanno qualche funzionalità marginale in più ma che costano circa 6 o 7 volte di più.

Torniamo al nostro obiettivo, che è la lettura di un file testuale Fixed Length. In particolare ci vogliamo porre l'obiettivo di realizzare una Macro che legga il nostro file testuale, che proponga in una cella la riga letta, e che chieda all'utente se caricare o meno quella riga.

In caso positivo la riga viene accodata in una zona del foglio, in caso negativo la riga letta viene ignorata e si passa alla successiva.

Figura 11 - Quattro Pro SE - Macro di lettura Riga per Riga. Nelle figure precedenti abbiamo visto come sia possibile caricare tutto un file testuale, e come una volta sul foglio, si possa manipolare, anche per ripartirlo correttamente nelle varie celle. In questo esercizio sperimentiamo un metodo del tutto contrario. Realizziamo una Macro, descritta nel testo, che legge riga per riga il file testuale. Propone la riga e poi chiede se caricarla o meno. In caso di risposta affermativa copia la riga in basso nella lista che appare in fondo.

Archivi	Modifica	Layout	Grafo	Stampa	Base	Varie	Opzioni	Finestre
A1: 'PROVA								
1	A	B	C	D				
2	PROVA	{(PONI C15;21){(PONI C16;""}						
3		{APRE "C:\TEST\PROVA.TXT";L}						
4		{PER C3;1;307;1;LEGGE}		54				
5		{CHIUDE}						
6		1199EVARISTO	GUALANDI	ROMA	M19610106A	2663000		
7								
8	LEGGE	{(LEGGERIGA B6}						
9		{ACCETTA "Premi un Tasto ";C9}						
10		{(ACCETTA "Caricare S/N/X ";C18}					X	
11		{(SE C18="S"){(SCRIVE						
12		{(SE C18="X"){(USCITA}						
13		{(RIENTRA}						
14								
15	SCRIVE	{(PONI C15;C15+1}		28				
16		{(PONI C16;{"A"&STRINGA(C15;0))}		A28				
17		{(PONI C17;{"B"&STRINGA(C15;0))}		B28				
18		{(PONI @CELLA("contenuto";C16);C3}						
19		{(PONI @CELLA("contenuto";C17);B6}						
20		{(RIENTRA}						
21								
22	14	1072UALERIO	AZZURRI	ROMA	S19710506A	3109000		
23	15	1078GIORDANO	GRIGIO	GENOVA	M19691206A	1471000		
24	16	1082ALDO	BORDEAUX	MILANO	M19750805G	5427000		
25	19	1090EVARISTO	CICLAMINO	MILANO	M19570117H	6560000		
26	20	1093GIIOVANNI	ARGENTO	PALESMO	S19540101C	6739000		
27	21	1098PAOLO	ORO	FIRENZE	S19510617D	5883000		
28	22	1102CARMELO	COBALTO	MILANO	M19721214E	8990000		
29								
30								
PROVA.WQ1	[1]	06/08/93	17:28	CAP				PRONTO

Il linguaggio Macro del Quattro Pro SE è analogo (le differenze sono marginali) a quello del Lotus 123, per cui possiamo consigliare la lettura di questa parte dell'articolo anche agli utenti 123.

La Macro (fig.11) contiene tre routine. La prima «PROVA» imposta le variabili (le celle C15, contatore di riga, e C16, stringa che contiene il riferimento della cella in cui posizionare la riga letta), apre il file testuale PROVA.TXT e attiva il contatore da 1 a 307 (sono tante le righe del file esterno).

Per 307 volte viene eseguita la routine LEGGE, che legge la riga e la posizione nella cella B6. L'istruzione ACCETTA serve come pausa, che permette di far leggere il contenuto della riga, e come INPUT, con il quale si sceglie se Accettare la riga, Scartarla oppure Interrompere la lettura.

Se la riga è accettata viene eseguita la routine SCRIVE.

Questa routine scrive due cose nella riga individuata dal contatore di riga, e che deve essere trasformato nel riferimento di cella. Scrive il numero della riga del file esterno (i buchi nella numerazione sono le righe scartate) e la riga stessa.

Non abbiamo manipolato, al contrario di quanto abbiamo fatto nella Macro di Excel, la riga. Si può scomporre facilmente con le istruzioni di sottostringa e di conversione.

Altra implementazione è quella che evita di dover scrivere il numero delle righe del file nella Macro. Occorrerebbe impostare un Loop «eterno» dal quale si esce se si intercetta la fine del file esterno. Ci può essere un tappo, oppure la fine si intercetta come situazione di errore.

Conclusioni

Insomma, se si dispone di un file con i dati, questo è in ogni caso trasportabile sul foglio elettronico. Si può leggere tutto insieme per poi ripartirlo nelle celle e manipolarlo con i normali strumenti del foglio, oppure, tramite una Macro da mettere a punto con un po' più di sforzo, leggerlo riga per riga e manipolarlo direttamente in fase di lettura. In questo caso il grosso del lavoro viene svolto dalla Macro, e si ha il vantaggio che i dati arrivano già nella quantità e nella forma voluta.

Un buon utilizzatore di spreadsheet deve, al momento opportuno, essere in grado di realizzare, in poco tempo, piccole Macro di utilità del tipo di quelle viste.