Welcome to Data Extractor

Extract any data and save it ready for use

Data Extractor can extract data from text. It is useful in all the cases where data is in various format and not structured inside a table or database.

As example you can have hundred or thousands of text emails coming from web form with labeled data as in the following example:

Name: John
Surname: Down
Address: 123 Spring St, Woodstock, IL, 60098, USA
Product: Red Box
Price: 35 US$
Date: 12 March 2012
IP: 123.456.123.456
Reseller: Brown Brothers Inc.

Copying and pasting these data from various files to a table to import them in a database can take a long time, specially if these data are buried inside text with various other data, often in a not so easy to read form, and if you have to do it not just for one file ... but for thousands and thousands of them.

It can take you weeks and weeks of work and maybe it would not be a such amazing job doing all this by hand, record by record.

Data Extractor come to the rescue

You can specify in the ‘Sources’ the folder containing all the file to parse and the extension of the files to parse.

In the ‘Accept Rules’ to accept any parsed file

In the ‘New Record Rules’ to create a new record for any file parsed

In the DataBase section you will create 8 new key fields

Any time you create a new key field a column is added to the internal DataBase columns.

You will end with a table with 8 columns (one for each different field)

In the ‘Extract Rules’ you will specify to extract data prefixed with the labels:

Name:
Surname:
Address:
Product:
and so on... and terminated by a newline character and you will specify to insert each of these data in a specific field you created in the DataBase section at the previous step.

Done.

You press ‘Extract’

The bar spins few seconds.

You can access the DataBase section and you will see the table was populated with all the data extracted and inserted correctly inside the database, record by record, each data in its own column.

Even if you had thousands of record to extract inside a folder, few seconds are enough to process all them.

Doing it manually can take a long time, days or weeks or months of work (if the data to parse and to collect is very large) that it is almost an impossible (or very expansive) job to do it.

Data Extractor can do it in few seconds. It’s magic!

Data Extractor can work on more complex situation of the one presented above, also where a single file includes more records, where more data fields are on the same columns, where the data to extract depends of it’s position inside a row of data, where files are in various place on the Hard disk, where fields have to be collected only if the data itself has a special prefix ... and so on and at any new release we add more and more extraction capabilities.

Data Extractor offers many option to let you specify what to extract

In case the feature you need it is not yet in this release, send us an email and say us what you need to extract data from your data set, we use users feedback to improve our products, release by release.
Getting Started with Data Extractor

Using the ‘Data Extractor Example’ documents

Data Extractor is bundled in the release we distribute from our site with a Demo Documents folder.
In case you purchased from the App Store you are welcome to our site to download the ‘Demo Documents for Data Extractor’ containing only the demo doc used here.
Use the Help menu in the application to come to our web site.

Download it and place it in your home folder where you prefer

Warning for the non expert:
Do NOT put the demo doc in your Application folder, use your application folder only for applications as any good Mac user has to do.

Open the Data Extractor document inside the ‘Lesson 1’ folder

Click in the upper tab bar and select the Sources tab if not yet selected

![Add button in toolbar](image)

Click the add button in the toolbar

The open dialog opens. Select the ‘Multi Files’ folder inside the ‘Lesson One’ folder.

![Selected folder in dialog](image)

Be sure that the .txt extension is in the accepted file types at the bottom of the window
Click the ‘Accept Rules’ tab
See that ‘Accept any files’ is selected as default if no rules are present.

Click the ‘New Record Rules’:

See that ‘Create a new record for any file’ is the default option

Click the ‘Extract Rules’ in the upper tab
See there are 6 rules, all very similar in their structure
We have a look at the first:

The first rule will get the data prefixed with ‘Quantity:’
and terminated by a newline
Any data will be accepted (regardless of the content)
If we examine one of the file inside the folder we selected in the source we can see
we have some 'Lore ipsum' latin text and some row as

Customer:  Liza Blanchard  
Address:  River Vista Avenue  
State:  
Transaction ID:0767453983287975  

And some random english text  

We mixed data and various text to let you see that Data Extractor is able to extract data even if data is mixed inside other various text  
It doesn't matter all the text around, Data Extractor can extract your data even if it's mixed inside a huge amount of other data and in any form, one record per document, many record in a document or one record per line.  
If you give Data Extractor enough informations it will extract the data you need in the correct way.  

Press now on the upper tab bar the ‘DataBase’ tab
You can see a table on the left and a table on the right. Focus your attention on the left one, the one containing the ‘DataBase Fields’

You can see there are 6 rows
At any time you can create rows in this table using the ‘+’ button.

The table on the right has 6 column (exactly as the rows in the left column)
For any row in the left column there is a column in the right table
The reverse is also true: any column in the right table (the DataBase) is represented by a row in the left table. Simply the left table is the list of the columns in the DataBase table.

Adding or removing elements in the left table (DataBase fields) adds and removes columns in the DataBase.
Go back to the ‘Extract Rules’ area
You can see (in the lower part of the window) that any extract rule points to a destination in the ‘Insert to destination’ popup

This is the way to control where the extracted data will be inserted in the database.
Press now the Run section in the upper tab bar
You can see the entire flow of the process is represented.

From the sources (just one folder)
to the found files (according to file types in case of folders)
to the accept rules (0 rules, default accept all)
to the rules to create new records (0 new record rules, default new record for any new file in action)
to the extract rules (6 rules)
to the insertion in the DataBase (with 6 fields available)

We are ready to go.
Press the ‘Extract from Sources’ button
Done
You see the info were changed
1 source was used, 20 files accepted, 20 records created and 120 fields inserted

Press the 'DataBase' tab bar
You can see the DataBase was populated with the extracted data as specified in the extract rules

Press the ‘Export to Text File…’ and save the file on disk.
You can select between the TSV or CSV format (tab separated or comma separated value)
The file can be imported in any database or app you like in this universal recognized format.
You have now an idea of the job Data Extractor can do for you.
To better understand how to use Data Extractor and see more example in action read the ‘Using More Examples’ and the ‘Data Extractor Reference’ section.
Using More Examples

Lesson Two

Open the ‘Lesson Two’ document in Data Extractor.
Add to the ‘Sources’ the only file included in the example, ‘One File.txt’
Press the viewer file at the bottom of the file list to visualise it in the internal viewer.

As you can see in this case we have a different situation then in the ‘Example One’ used in ‘Getting Started with Data Extractor’

Here we have a single file containing more records, one by line, in a multiline text

We can examine the first line:

Subject: Thanks for your purchase 
Quantity: 5; Product: Buccellato; Wine included: Gavi o Cortese di Gavi; Customer: Deann Mosley; Address: Glendale Road; Transaction ID: 1943507167328740;

Subject: Thanks for your purchase 
Quantity: 1; Product: Devil’s food cake; Wine included: Chianti; Customer: Ricky Dickson; Address: Mcelroy Drive; Transaction ID: 5295649401299802;

Switch to the ‘Accept Rules’ area: It uses the default set, it’s ok.

Switch to the ‘New Record Rules’ area: It uses one rule to create a new record at ‘newline or Return’. This to create a new record per parsed line
The ‘Extract Rules’ now: It contains 6 rules, all with a starting tag, any data accepted and all using as data suffix ‘;’.

Any extraction rule points to a different field in the database.

Look at the first examined row of the ‘One File.txt’ file and at the first Extraction Rule you can see it will extract the quantity of ‘5’ which is prefixed by the string ‘Quantity:’ and with a suffix ‘;’.

Quantity:5;

In the second line quantity is ‘1’ recognized by the same prefix ‘Quantity:’ and suffix ‘;’.

Data extractor can recognize the field by their prefix and suffix.

Now switch to the Data Base, you can see the 6 fields used as destination by our ‘Extraction Rules’. (they were created in advance, before using them as destination for the ‘Extraction Rules’).

Now switch to the ‘Run’ tab area.

The status says ‘OK’ It does mean we have no error in our documents.

Press the ‘Extract from Sources’ button.

You can see: 19 create records and 114 inserted fields.

Switch back to the Data Base.
You can see the 19 created records ready to be exported on disk.

![Data Extractor Interface](image)

**Lesson Tree**

Open the ‘Lesson Tree’ example
Add in the ‘Sources’ the file ‘allInOne.txt’
Open in in the viewer pressing the ‘Open in the viewer’ button

In this case we have a single file where multiple records are spanned over a multiline text with a lot of various text noise inside

In that case we need to instruct Data Extractor when to create a new record
Analyzing how the layout is organized we recognize that a ‘Subject: Thanks for your purchase’ identify the start of a new data block in the text

So we created a ‘New Record’ rule that create a new record anytime this string is found
Examining the Extract Rules you will see a similar structure of the previous example. The records are just newline terminated and the rules contain it as suffix.
Extract and visualize the results.

**Lesson Four**

Open the ‘Lesson Four’ example
Add the ‘Multi Files’ folder inside the ‘Lesson 4’ folder
The structure of the document is similar to the ‘Lesson One’
Go to the run section and run the ‘Extract from Source’
See the result

If you open one of the files inside the example
Using the ‘Open from Disk in Internal Viewer…’

you see the result have something wrong

The Quantity value is wrong because in the text file to extract a tag ‘Quantity: ‘ is present more then one time before the true ‘quantity’ value we wanted to extract
We can see examining all the files that the true data block has a string at the start saying: ‘Subject: Thanks for your purchase’
We can use it to instruct Data Creator to start extracting only after a specified string to avoid all the entry before.

To solve this problem we can modify the extract rule for the Quantity and insert a ‘Search from text:’ with a value of ‘Subject: Thanks for your purchase’
The app will start to search for any file at the correct place skipping all the data before.

We can now reset the extracted data using the ‘Reset extracted data…’ and extract again.

Now we will have the correct results.
Data Extractor Reference

About the app

Data Extractor can extract data from text. It is useful in all the cases where data is in various format, identified by labels and not structured inside a table or database.

Data Extractor, in one click only, can:

- Parse text files on your Hard Disk
- Using a custom extraction workflow extract all the data you need
- Insert them in an internal database table ready to be exported in TSV (tab separated value) or CSV (comma separated value) format

All the workflow can be saved on disk and re-used at any time with a single click.

Data Extractor allows to extract hundred or thousand of record in few seconds.

Doing manually what Data Extractor can do is time expansive if not impossible due to the complex and time tedious operation involved if performed manually.

Data Extractor is a ‘Sandboxed’ app conforming with the latest Apple requirements.

Data Extractor is ‘Sandboxed’ both in the ‘App Store’ release and in the ‘Open distribution’ release we have to download at our site.

Both are also digitally signed with an original developer certificate Apple released to us to provide the new level of security the new Apple technology offers to the users.

Document based

Data Extractor is document based application. Inside any document all the settings used for extraction can be saved and retrieved at later time. In that way it is enough to have a separate document for any kind of extraction to perform, and reopen it and launch the extraction with a simple click.

Organisation of a document

The main Data Extractor document has an upper tab bar that allow to navigate in all the six areas the work is divided.

The ‘Sources’ area

The Source area allows to set the sources (files and folder) to parse for documents to scan to extract data. Folders will be parsed for files at any deep level. Files will be opened only if they match with the ‘Accepted file type’ field at the bottom of the window. Any insertion has it’s own ‘Accepted file types’
Files inside the folder not matching with any of the accepted file type will not be scanned for extraction. If you don’t obtain result scanning a folder ask yourself if you added the file type of the files you need to scan.

When a folder is added the initial ‘Find file types’ are the one set in the preferences . To add more modify them in the preferences

To add a file or a folder to the list:
Drag and drop folders of files from the finder. Done

or
Press the plus button in the toolbar
or
Press the plus button at the bottom of the source list

An open dialog will open
You can select file and folders
Multiple selections are allowed

Files and folder can also added using drag and drop from the finder to the source table.

To delete a folder or a file from the list select them and press the minus button from the toolbar
Or press the minus button at the bottom of the list
When a source is deleted from the list, it is deleted only from the list
The original is ALWAYS leaved untouched on disk.
To say more, Data Extractor ALWAYS open file to scan in read only mode from the sandbox

Considering the operations are performed from the sandbox, It has no way to damage, overwrite or delete the files you insert in the source list when you manipulate them in the sources list.

You can undo the add and delete file/folder operations

When a file is selected you can view it inside the internal viewer Data Extractor provides.
It is an easy way to select and copy tag string to use in the ‘Extract Rules’ section
The internal Viewer open files in read only mode. It can’t write on the files you view.

You can also view files not listed in the sources using the internal Viewer. Select from the tool popup at the bottom of the source list the ‘Open File from Disk in Internal Viewer…’ commend and select from the dialog the file.
It will be opened in the internal viewer.

A file to be parsed need also that the ‘Use’ check box is selected

You can select or unselect all the sources in the list to be use enabled from the popup tool

Think at these option as a filter
If a file is selected for use and in case of a file inside a folder it's extension matches with the ‘Accept File Type’ it passes the filter and is sent to the successive area, the ‘Accept Rules’ area

The ‘Accept Rules’ area

Accept rules work on all the files that are accepted for parsing in the ‘Sources’ area
Think at it as a second filter before reaching extraction
If the list of the accept rules is empty, the default ‘accept any file’ is in action and any file coming from the source list is accepted and sent for extraction
Adding an ‘Accept Rule’ allows to select different options:

- Acceptance based on the file name (extension not considered, if you want to accept files based on the extension use the feature in the ‘Source’ section).
- Acceptance based on content of the file (content of file considered in its full extent) (optionally case sensitive).

It is possible to accept a source file if any of the option is satisfied or only if all the listed condition are satisfied.
Be careful with this last option (All listed condition satisfied) because it is easy to set conditions where no file satisfy ALL of them.

The ‘New Record Rules’ area
Files that pass the first (sources) and the second filter (accept Rules) come to the ‘New Record Rules’

When Data Extractor extracts data it insert them inside a record. A record is represented by a row in the internal DataBase. Data extractor needs to decide when to stop insert data in the current row and create a new one where to insert new extracted data.

Doing it when all the row is full it would be not a good idea simply because a record could, as example, miss a field and the app will not create a new record ... and it will overwrite the previous collected row with the new data of a new parsed record.

Ouch!

If you have to extract data where not all the fields are always present in records, the app will end to collect nothing expecting to find ‘complete’ records it will never find.

A more rational approach is to let the user decide when the app has to create a new record in the database. The ‘New Record Rules’ allows to do it!

You can specify more then one condition. Any of this condition trigs the creation of a new record (if there are no successive data to insert, the last record will not be created)

First type of trigger is when a new file is parsed. This is a very common situation where a single record is not distributed in more then one parsed file, so parsing a new file means a new record can be created to extract data to (if it founds something to extract). Very simple concept.

Second trigger is newline or return. This is a case where any single record is on a single line of text. When the app parse a new line of text it knows it can create a new record (if it founds something to extract)

Third trigger is more generic. It is possible to specify a text entry in a field (optionally case sensitive). When the app encounters the specified string it create a new record (only if it founds something to extract that follows, to avoid creation of an empty record).
The ‘Extract Rules’ area

The ‘Extract Rules’ area is the most important in terms of data elaboration inside the app.

If you won’t learn how to use it well, you will never be able to have Data Extractor working for you in the proper way.

It allows to pull data from any ‘record block’ of data passed from the 3 previous filters and insert in a smart way these data in the correct fields in the internal database.

The extract rules are created to allow each rule to extract one specific data and insert it into a specific field. You can create more than one rule pointing to the same destination field.

You can specify for any extract rule:

- A name (not used in the process, just a label useful for yourself to recognise all your rules)

- A ‘Search from begin’ or ‘Search from text’ to indicate where to start searching the data to skip previous misleading insertion and avoid to collect unwanted data
  - as example in emails there is a ‘Date’ field in the header, you may need to skip this one to get your true ‘Date’ field

- A ‘Data prefix’ which is the label immediately precedent to the data to extract (optionally case sensitive)

You can select to have data prefix valid only if at start of line.
You may also indicate for the prefix an optional ‘NewLine at end’ to indicate the data to be extracted is on a new line and the prefix has a newline at its own end.

Then the data filter to indicate what to get or not. A data filter that can be:
• any data
• data starting with a specific string
• data ending with a specific

In case the data filter is not satisfied the data is not extracted.

Or you can apply a sort of post-processing the the extracted data and sub-extracting a specific portion at a specific position selecting ‘Extract at data position’.

inside the block specified by the ‘Data prefix and the ‘End of Data’

See successive paragraph ‘Sub-Extractions of values in detail’ for more about that.

Then you have to specify the end of the data portion to extract.

- terminated by a newline
- terminated by a tab
- terminated by a string or a single char (optionally case sensitive)
And you need to specify where to insert the data extracted selecting the field (in the DataBase) where to insert the extracted data

Specifying a destination field is mandatory!

More then one rule can point to the same field, in case of multiple write the last will overwrite the previous

While inserting and editing extraction rules you can create new fields without switching to the 'DataBase' section directly from the 'Extract Rules' area using the 'New Field' button

This is useful creating new extraction rules if not all the fields were created in advance

**Sub-Extractions of values in details**

Release 1.4 and successive introduces (at popular demand) sub-extraction of values

You can extract something inside the block specified by the 'Data prefix and the 'End of Data' based on it's position inside the data

In same way you specify a sub-extraction on the previous extraction, based on position and not based on prefix and postfix labels

As example extracting
form the text:

name: Jane Smith
Date: January 26, 2015
Country: USA
Values: 121  317  806  5592
Status: active

always the 2nd value of the row ‘Values ‘(which in the example is ‘317’) where in the row the four numbers are separated by tabs

You will indicate

‘Values.’ as start data

and you will indicate ‘New Line’ as end of data

this will let you get the entire row: 121  317  806  5592

You need further to specify:
‘Extract data at position’ inside the ‘Accept Data’ block

then using the submenu ‘Get Item inside block as:’

and you will specify method to use and position to extract

Then Data Extractor will get what specified by the ‘start’ and ‘end’ of data and will pick the 2nd position in the row
The auto detect feature some time is smart enough to can in any case if able to auto-recognise the used method in the data block.

The ‘DataBase’ area

The database area specify the fields (columns) of the database (table) where the app will automatically insert data.

Data that passes the previous filter and is extracted by the ‘Extract Rules’ are put in the ‘DataBase’ which simply is an internal table with a custom number of columns and an unlimited number of available rows to accept and hold the extracted data.

During extraction the database (table) will be filled with the extracted data.

The database must be set before inserting any data (you create the columns in advance adding keys to the key list on the left)
To insert and delete columns in the database use the plus and minus button at the bottom of the key list.

For any new key list, a column is added in the Database table.

There is a correspondence one to one from keys and columns in the table.
You add a Key in the database and a column is added.
You delete a key and the correspondent column is deleted from the table.

To change the name of a key field double click it and edit it in-place.

You can move database keys (and the associated column in the database) using the up arrow and down arrow.

The app keeps always the keys and column position in sync.

After the extraction, the database will be auto-filled with the result of the extraction.
Records (rows in the database table) can be selected and copied using the utility popup at the bottom of the table.
Records (rows) can be selected and deleted

All the data inside the table can be deleted using the ‘Reset Extracted Data’ command

To export data use the ‘Export to text File’ command
The export dialog allows to use the fields name as the first exported row and select between the TSV (tab separated value), CSV (comma separated value with comma) and CSV (comma separated value no comma) and format to use for exporting your data.

You can also select from the table rows (also not contiguous selection using the command key) and copy to the clipboard the selection in the format you like from the action popup.
The ‘Run’ area

The ‘Run’ area show a report of all the previous areas and allows to launch the extraction

You can see the report of the extraction with details regarding the single steps

Before starting the extraction, if something in the setting is wrong, the app informs you
You can go in the area specified by the error message and fix it. In the panel the error will be highlighted too.

The error message disappears as soon as the error is fixed.

**Drag and Drop processing**

The Drag & Drop area allows to process a file for extraction as soon as you drop on it a file from the finder.
See the section ‘Using the Example’ to see a use of Data Extractor with different kind of file to parse

**Modern MacOS features**

Data Extractor adopts and makes use of the latest Apple technologies

They are used in the standard Apple way, so nothing new to learn if you already know how to use them.

- **Resume** – the app will reopen at the point and state the user left it included opened documents and unsaved one
- **Auto save** – the app saves using the Mac OS X Lion autosave functionality
- **Versions** – the user can look at previous versions of documents and restore to any earlier version (just select Revert to Saved from the menu or the down arrow at the right of the title window, it appears only if you move the cursor over)
To use the Versions you can select ‘File > Revert To > Browse All Versions’ from the menu

- Full Screen - the user can switch at any time to full screen using the native OS X Lion full screen mode and commands (use the upper right icon in the window to go full screen and bang the mouse in the upper part of the screen to re-obtain the menu and eventually clicking the standard icon to exit the full screen mode)

Dark Appearance Mode

Data Extractor is fully compatible with Dark Appearance Mode available on macOS from macOS 10.14 Mojave and obviously with Light Appearance Mode on any macOS release it can run on.

Preferences

General

Specify the action to do at startup:

- Standard Behaviour
- Open Dialog if No resume
- Open Dialog Always
- Open Selected document
To select a document press the select from disk button and select it from the open dialog that will follow.

At any successive relaunch Data Extractor will execute the option selected. The additional 3 buttons at the foot of the edit field let you:

- Show in the finder the selected file
- Test open the selected file as it will be done at the next application launch
- Delete the bookmark reference to the file (you will need to reselect a file)

The reference to the file is taken using scooped bookmarks and they work even if you move the file on your disk, even if your file is in the trash.

To update the path display in case the file was moved, click the ‘show in finder’ or the ‘test open’ button.

**Advanced**

- Allows to insert the file extension inserted as default value when you add a new folder in the ‘Sources’ area
- Allows to reset the counter dragging a file on a drag & drop area
- Allows to reset a counter launching a new extraction

Update

* Update section in the preferences is not available in the App Store Release (if you purchased via the App Store you don’t need it, to obtain an update use the App Store Update function)

Data Extractor can inform you if an update is available. We suggest to keep this option on.

When you check if an update is available, the application ping the Tension Software server and receive an answer about the last release available, the application can in that way know if its obsolete or not. In case the application will ask you if you like to download an upgrade.

If you download the upgrade, you need to install it.
A common error is to download an upgrade and install it maintaining the old application somewhere on the hard disk. Then using to open the documents sometime the old application and sometime the new one. This cause some problems*. To avoid it simply install your application in the place dedicated to it, the Application folder. When you install an update on the standard Application folder, the system will ask if you want to replace the old one with the new, answering yes will install the new one replacing the old.

*If the application says the document you are trying to open was created with a newer version of the application, probably you have two different release of the application on your Hard Disk and you are trying to open a document with the older version after having modified the document with the new one. When you receive a similar message open the about box inside the application and see if you are running the last release. In case download it, install it and USE it!

Help

Data Extractor provides a user guide in PDF.

Support

You can also obtain support using the ‘Support Email...’ command. An email will be prepared using your email client with the correct address to send to. Yes, we answer to your emails.

Data Extractor is a Commercial Program

Data Extractor is available via the Apple App Store (fully enabled) and via our site (in an unlicensed version)

In case of the App Store release, a license is already included with your App Store purchase and you don’t need to buy a license

If download from our site or on the web elsewhere, you can use our software for a test period of 10 days
After that you are required to buy a license to fully enable it and be legally authorized to continue to use our software

Licensing Data Extractor

If you are using a version from the Apple App Store: a license is already included with your App Store purchase and you don’t need to buy a license

If you are using a version download from websites: you can buy a license to use our software using the ‘Buy License’ command in the ‘License’ dialog.
You will open in your browser our license web page on our web site
From there you can buy a license to use Data Extractor using one of the payment service we provide. It easy, fast, and secure and all most important form of payment are accepted.
We carefully selected or international reseller to be sure the buying experience for our customer will be as best as possible and without any problems.
After purchase you will receive from Tension Software a license email containing your full name and a license code
Open the License dialog with the command ‘license’ under the ‘Data Extractor’ menu and insert these data in the License dialog. The software will become fully licensed and fully enabled for future use.

Name and license code are remembered by the program and don’t need to be re-inserted at successive launch.
In case you move on a new Mac you need to re-insert your name and license
The license is valid for a single Macintosh. You can purchase for two Mac as long as they are not used at the same time, as example a desktop and a laptop used by the same user.

About Tension Software and Ecleti

All the software asset by Tension Software is now under Ecleti
New company name but still the same owner and main developer (Roberto Panetta)
So for the users nothing changes. We provide the same software and the services.
All the user licenses are valid and assistance is provided as before to all our users.