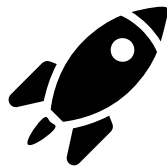


Introducing eSearch Pro

First public beta
30-day trial launch



30 September 2024

eSearch Pro 1.0 is a work in progress and still in the beta test phase.
Some functionality may be limited.
2024-12-20 Issue 9

Introduction

eSearch Pro is a comprehensive full-text search application with the unique ability to be translated by you.

German and French sample language files are supplied, along with a free Language File Translator, allowing you to quickly edit these or other partially translated language files.

eSearch Pro is being developed to be a useful alternative to dtSearch Desktop for our European customers, hopefully it will also be of interest worldwide.

We have been distributing dtSearch products in Europe for 30 years, but it is only available in English. Recent versions of dtSearch use our multilingual stemming files, but licensing restrictions mean we cannot offer a translated version.

eSearch Pro lacks some of the eForensic features of dtSearch but we aim to provide functionality and new features that dtSearch and others lack.

These differences and others make it possible to use eSearch Pro for applications that dtSearch or others may be less suitable for, we have presented a few ideas here...

Contents

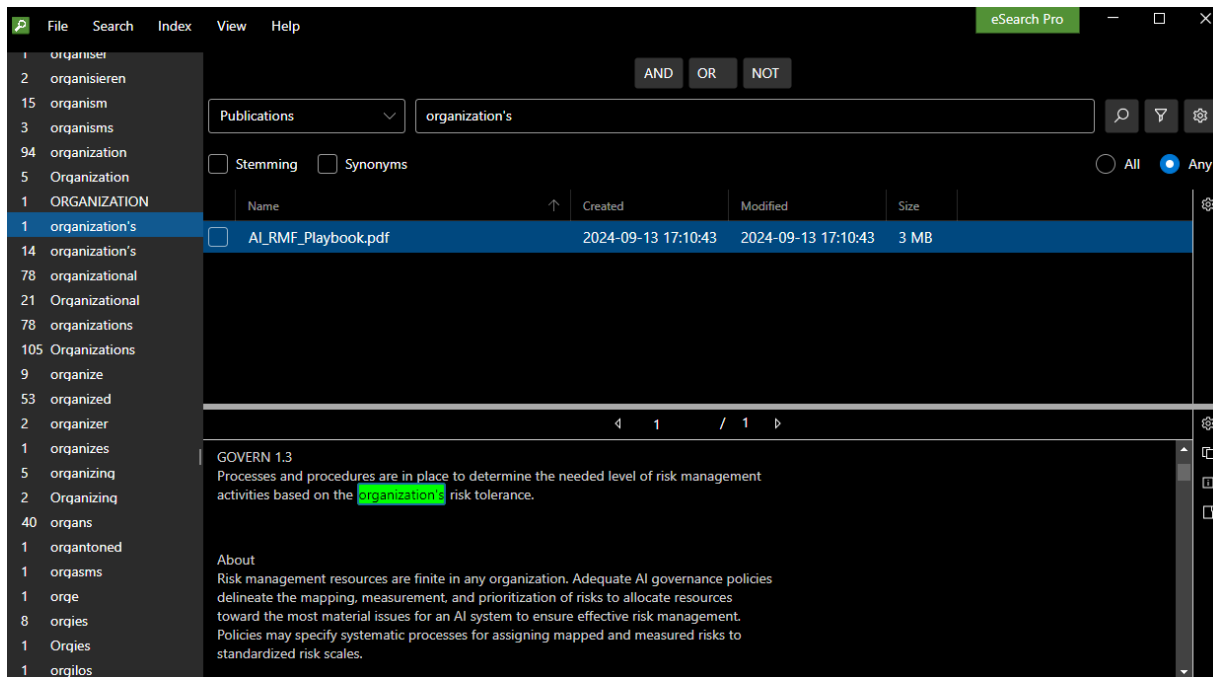
Introduction.....	2
Viva la difference!.....	3
Proof Reading.....	4
Geolocation in images.....	5
Locate Numeric Patterns.....	6
Search for Email addresses.....	6
Translation.....	7
Software Engineering.....	8
Large CSV Datasets.....	9
Example - ICIJ files.....	9
Example - TREC 2.....	10
Research or Publish Newsletters and Notebooks.....	11
30-day trial now available!.....	12

Viva la difference!

Here are a few notable differences between eSearch Pro and dtSearch:

- A valuable feature that dtSearch has, but others lack is a word list of all the words in the index. eSearch Pro takes feature several steps further:
 - Full length fully scrollable word list on the main window.
 - Sorted alphabetically for accented characters & uppercase letters.
 - Retains punctuation within numbers, the possessive apostrophe and contractions.
- Image formats (such as .png .jpg .gif .bmp .svg .tif) are displayed in the viewer; the viewer can also display a marker at the geolocation coordinates of an image on a built-in map view.
- Options for ‘Search as you Type’ and ‘List Index Contents on Empty Query’ (You can add words to limit the results to just the ones you need).
- Viewer plays browser compatible audio and video file formats like .WebM, .Ogg, .Wav, MP3. Just click ‘open in folder’ to play other formats in their default application or the recommended app.
- Source code can be searched and displayed with syntax highlighting and line numbers, 40+ source code file types supported.
- A filetype tree-view selector makes it easy for you to limit your index contents to just the filetypes of interest without having to move the files.
- ePub, Jupyter Notebook, Markdown file support.
- Portable version is self-contained and will run on a USB flash drive. You can index & search files stored on the host machine’s hard drive or a portable library on the USB flash drive itself for your own use or to publish search-only copies to others.
- Search results include not only a file’s Modified date and time, but also the Created and last Accessed date and time, all in ISO 8601 format (YYYY-MM-DD).
- Hit navigation includes current hit number as well as total count.
- Did we mention dark mode as default?

Proof Reading

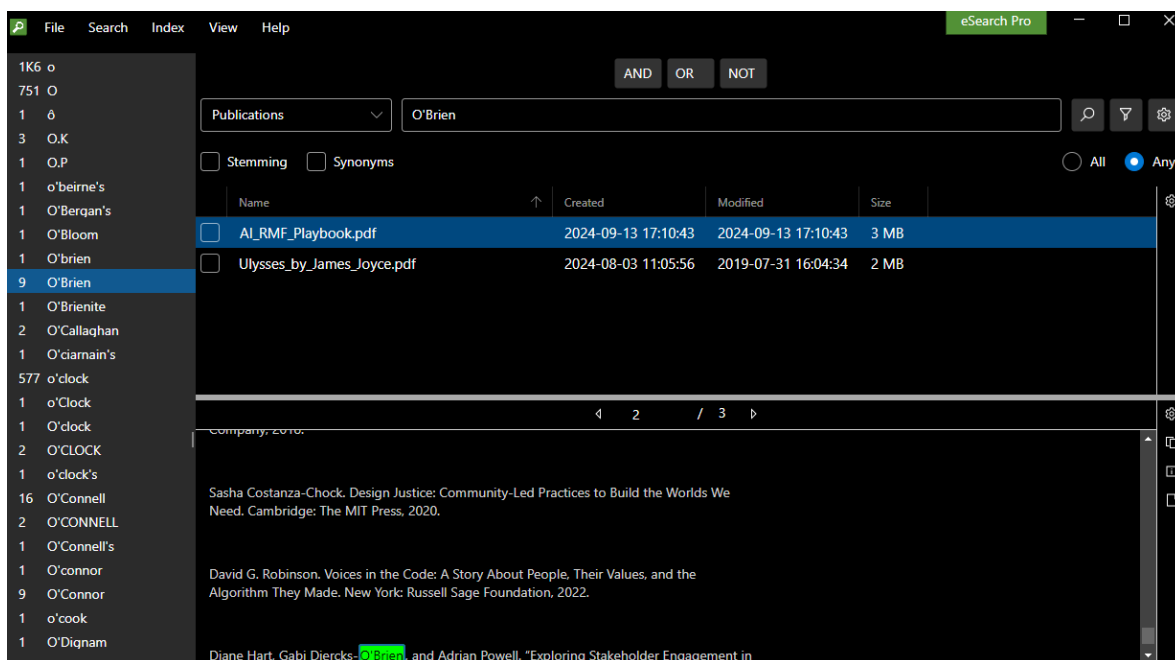


Whether you are a professional proofreader, a translator or a just a nitpicker, you will love the scrollable word list that shows every word in the documents, just double-click on a word to highlight them.

The default indexing method is 'accent sensitive' so that cafe doesn't pass for café.

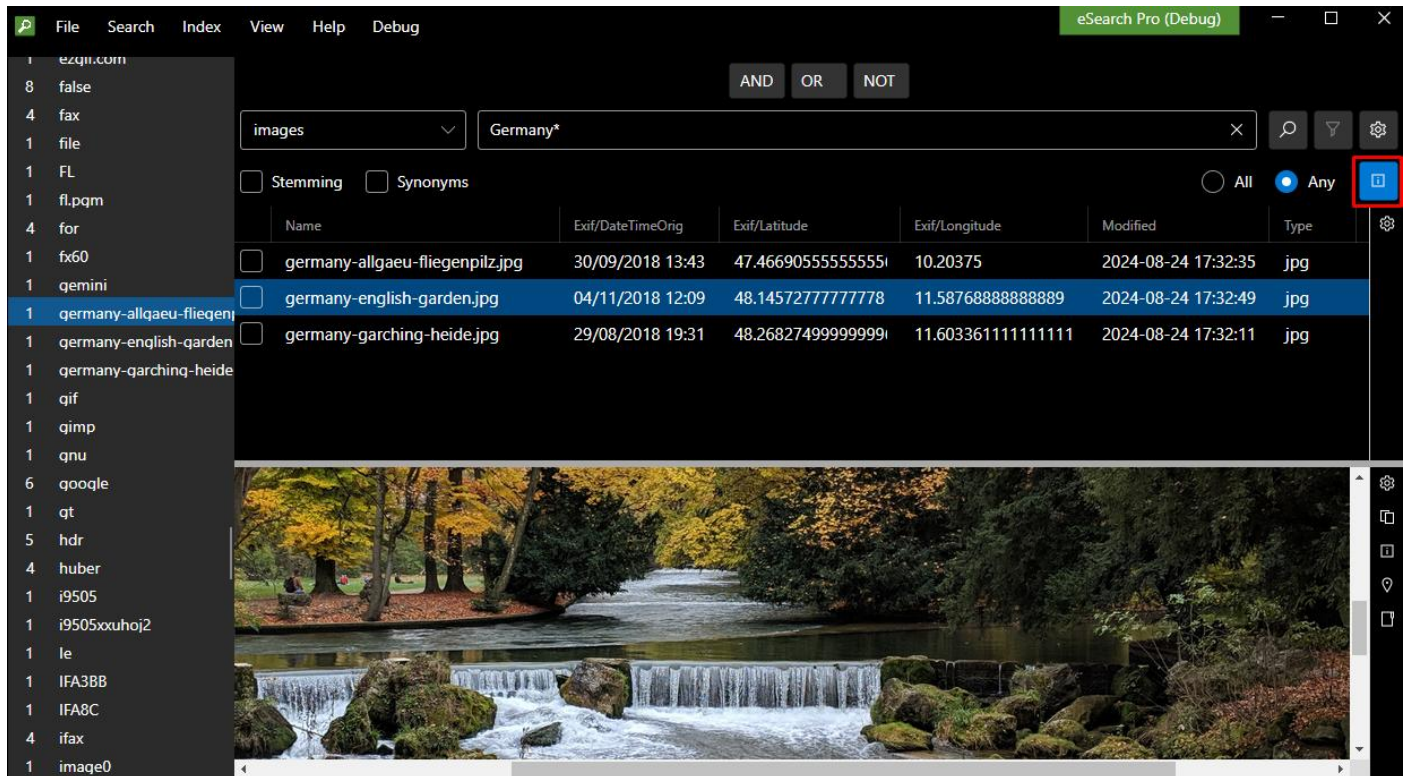
Case Sensitive indexing option ensures that references to a Polish delicacy doesn't get mistaken for a kind of wax polish.

Indexing preserves punctuation within numbers and words like dates, IP addresses, library formats, sums of money, or abbreviations. Possessive apostrophes with different symbols (U+0027 or U+2019 as used in the NIST AI RMF Playbook shown above) are easily spotted.

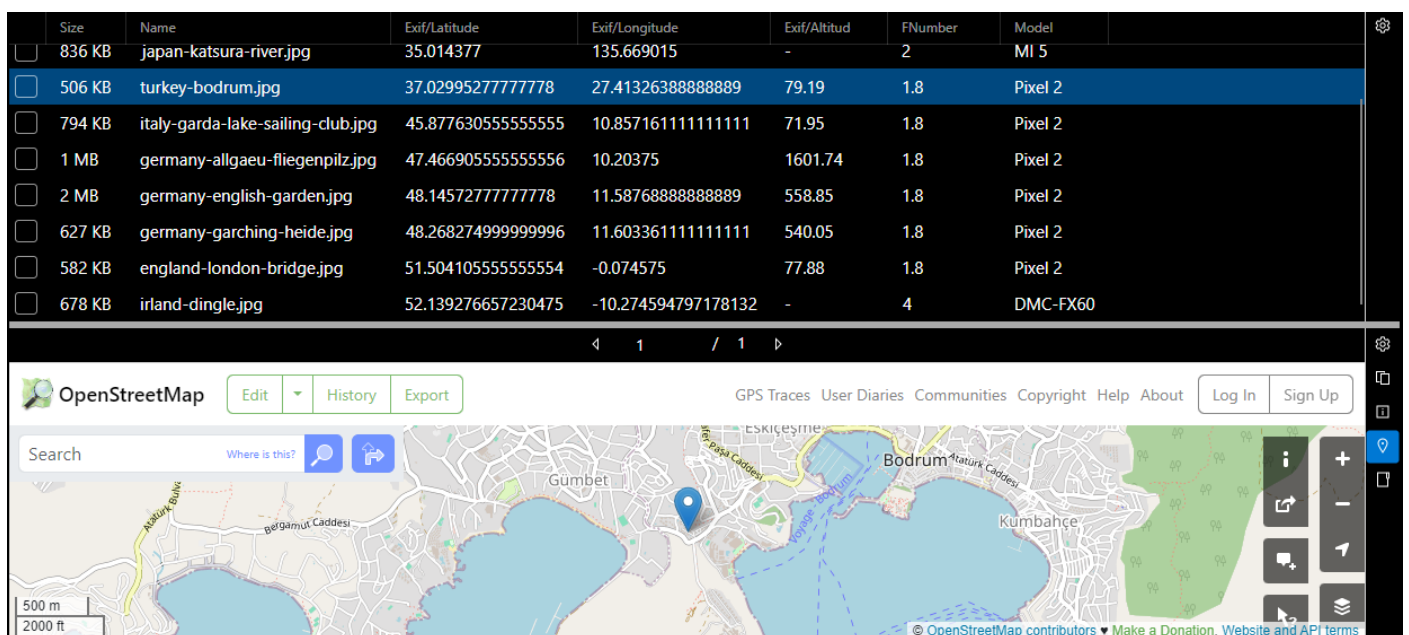


Geolocation in images


eSearch Pro displays images and automatically extracts metadata such as geolocation, camera settings and more. You can choose the specific types of metadata you want to see in your results columns. Additionally, you can click and drag header columns to reorder them and sort the results based on the selected metadata key.



You can also choose to display the geolocation on a map!



Displays images and lists metadata in these formats: .bmp .gif .jpg .pgm .png .svg .tiff.

 TIP: Keep images in separate indexes and press the Search Metadata button to see just the image metadata in the wordlist.

Locate Numeric Patterns

eSearch indexes numbers by default and preserves punctuation within numbers. This makes it possible to search for patterns within them using Regular Expressions.

eSearch includes an example regular expression in the Sample Numeric Patterns synonyms files to identify IPv4 internet addresses. You can also add your own regular expressions, it uses the same Lucene syntax as Microsoft Azure AI Search.

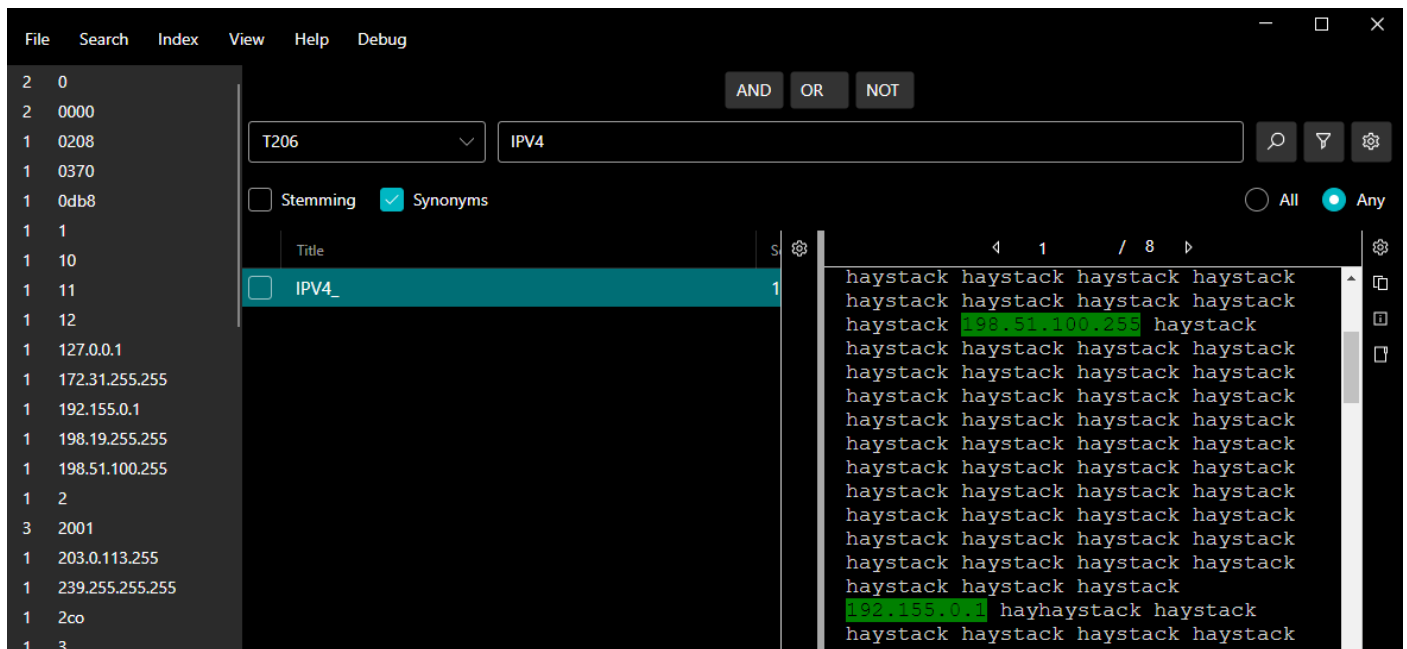


Image above shows IPV4 search query used to find and highlight all IPv4 addresses in a file.

Search for Email addresses

Email addresses are split on the @ symbol so that domains and email aliases appear separately in the word list.

You can search for all users at contoso.com just by entering that domain as the search query (do NOT use a wildcard) or you can enter the alias jon.doe to find all such emails at any domain e.g. jon.doe@gmail.com or jon.doe@contoso.com

To search for the exact email use both parts in quotes including the @ sign like "jon.doe@contoso.com".

Unlike dtSearch there is no need to specify an option in the index settings to index emails, it is the default behaviour.

Translation

eSearch Pro supports .xliff (1.2 & 2.0), .tbx and .tbm file formats used by translation software. The index is accent sensitive by default and Index Settings offers a Case Sensitive option, so that for example Language, language and LANGUAGE get indexed and appear in the word list as three separate words.

Multilingual stemming means that plurals and other word variations are found in many European languages.

Pre-defined synonym files are included that cover irregular verbs and nouns in Danish, Dutch, English, French, German, Italian, Norwegian, Spanish, and Swedish. You can choose to use several synonym files during a search. Synonym files also include multilingual days and months.

The word list displays contractions like d'un and possessive apostrophes.

The screenshot shows the eSearch Pro interface. On the left is a word list with 'd'un' selected. The main area shows search results for 'd'un' in a file named 'sample-1-xliff2.xliff'. The results are displayed in a monospaced font.

Name	Created	Modified	Size
sample-1-xliff2.xliff	2024-09-04 16:43:12	2024-09-04 16:43:12	656 B

Search results for 'd'un':

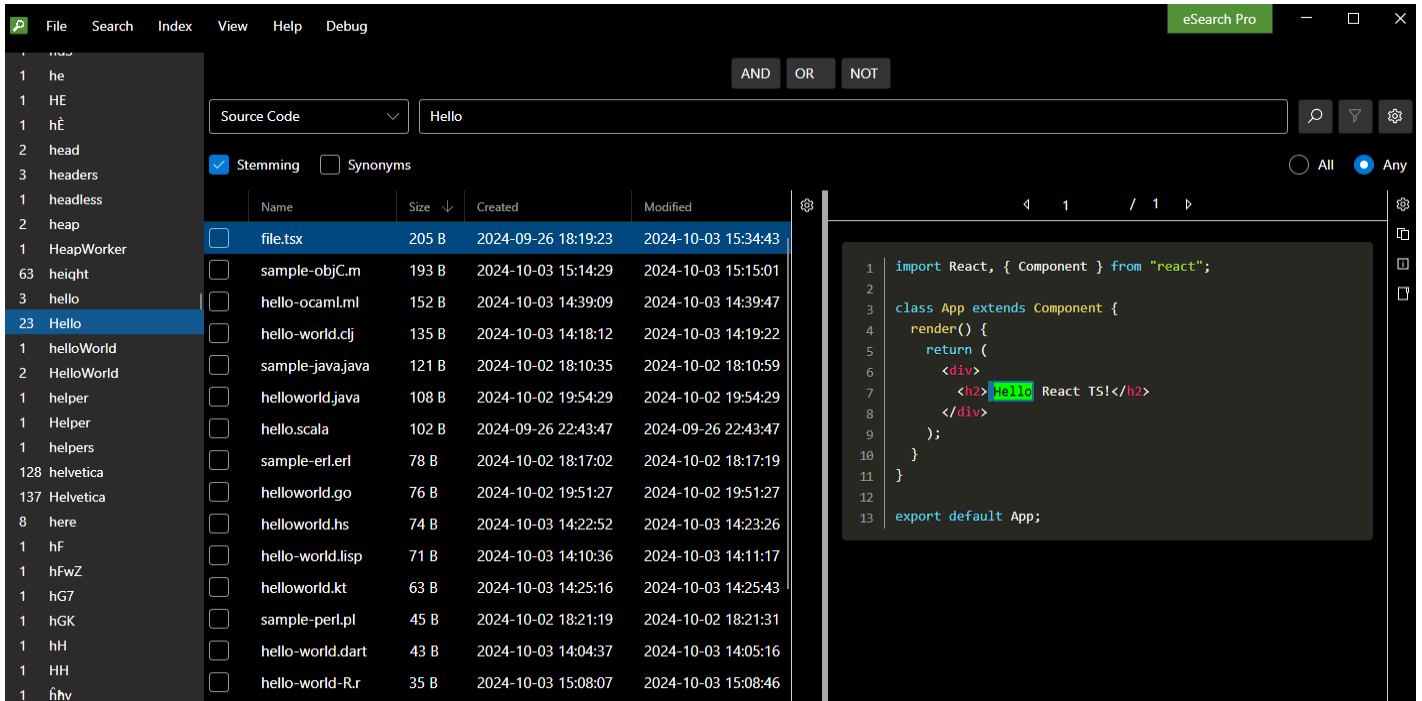
```
Hello world  
  
Bonjour le monde  
  
This is an XLIFF document.  
  
Il s'agit d'un document XLIFF.
```

*As of 1.0.9120 (beta) release it is only suitable for small files, eSearch Pro currently does not have a way to segment large files.

The free GSsplit 3.0 by G.D.G. Software SARL can be used for this purpose. <https://www.gdgsoft.com/gsplit> The UI is available in many languages.

Software Engineering

Over 40 source code file formats are supported, the viewer displays source code with hit highlighting, syntax highlighting & line numbers.



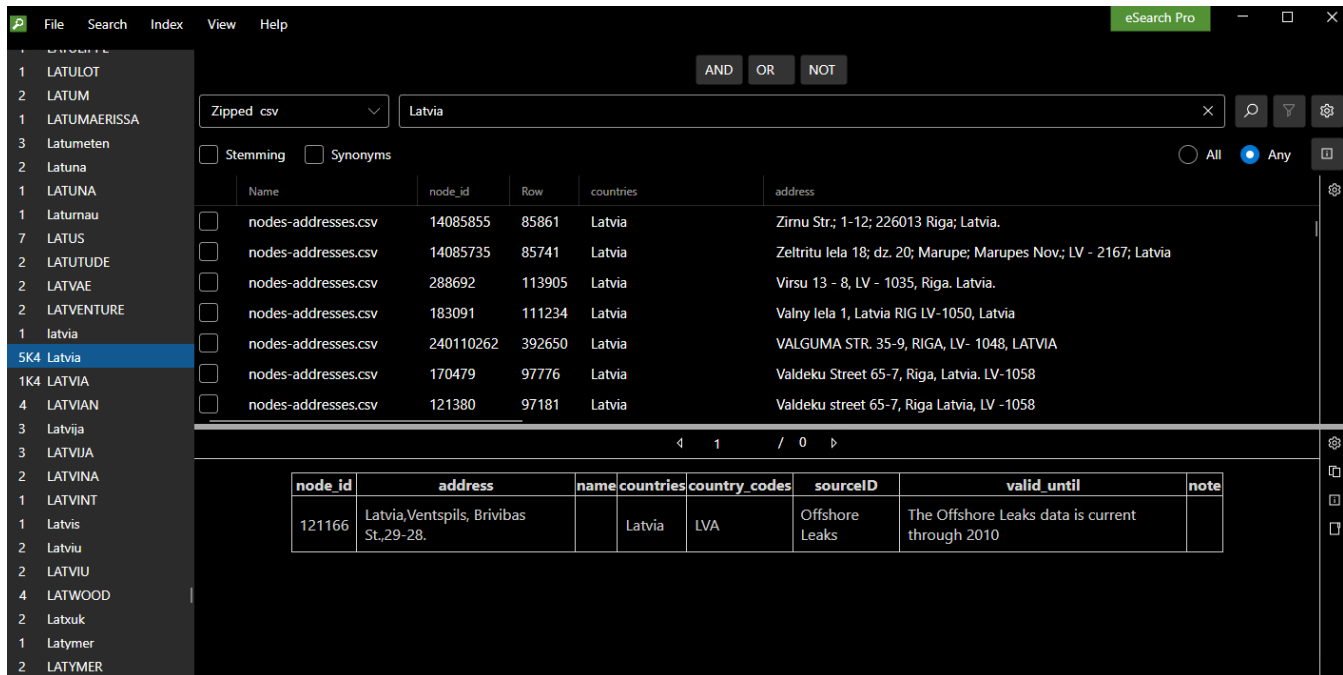
Example shows Created and Modified date & time all in ISO 8601 format (YYYY-MM-DD), file size in Bytes and filename.

eSearch Pro 1.0.9120 (beta) supports .asm .bat .c .clj .coffee .cpp .cs .dart .elm .erl .fs .go .groovy .hs .java .jl .js .jsx .kt .lisp .lua .m .ml .php .pl .pro .ps1 .py .r .rb .rkt .rs .sass .scala .scss .sh .sql .swift .ts .tsx .v .vb .vhd

Large CSV Datasets

eSearch Pro can index .csv files inside an archive such as Zip. It will automatically index all metadata and display each record separately. You can select which columns to display in the search results.

Example - ICIJ files.



The screenshot shows the eSearch Pro interface with a search for 'Latvia' in a 'Zipped csv' file. The search results are displayed in a table with the following columns: Name, node_id, Row, countries, and address. The results are as follows:

Name	node_id	Row	countries	address
nodes-addresses.csv	14085855	85861	Latvia	Zirnu Str; 1-12; 226013 Riga; Latvia.
nodes-addresses.csv	14085735	85741	Latvia	Zeltritu Iela 18; dz. 20; Marupe; Marupes Nov; LV - 2167; Latvia
nodes-addresses.csv	288692	113905	Latvia	Virsu 13 - 8, LV - 1035, Riga. Latvia.
nodes-addresses.csv	183091	111234	Latvia	Valny Iela 1, Latvia RIG LV-1050, Latvia
nodes-addresses.csv	240110262	392650	Latvia	VALGUMA STR. 35-9, RIGA, LV- 1048, LATVIA
nodes-addresses.csv	170479	97776	Latvia	Valdeku Street 65-7, Riga, Latvia. LV-1058
nodes-addresses.csv	121380	97181	Latvia	Valdeku street 65-7, Riga Latvia, LV -1058

Below the main table, a detailed view of a record is shown with the following columns: node_id, address, name, countries, country_codes, sourceID, valid_until, and note. The record for node_id 121166 is as follows:

node_id	address	name	countries	country_codes	sourceID	valid_until	note
121166	Latvia,Ventspils, Brivibas St.,29-28.		Latvia	LVA	Offshore Leaks	The Offshore Leaks data is current through 2010	

Example above from International Consortium of Investigative Journalists Offshore Leaks zip file. <https://offshoreleaks.icij.org/pages/database>

The 70MB 'full-olddbLATEST.zip' file includes Offshore Leaks (2013) Panama Papers (2016) Bahamas Leaks (2016) Paradise Papers (2017) Pandora Papers (2021), a total of 5,356,936 records. On a Windows 11 Pro Intel i7 32Gb RAM machine a case sensitive index of that zip file with no stop words removed takes approximately 4 minutes to index.

Example - TREC 2.

The screenshot shows the eSearch Pro interface with the following details:

- Search query: Musk AND battery
- Dataset: TREC 2
- Filters: Stemming (unchecked), Synonyms (unchecked), All (selected), Any (deselected)
- Results table:

Name	Row	msg_dt	message
trc2.csv	1588416	2009-01-13 17:55:50	AUTOSHOW-UPDATE 1-Tesla to supply battery for electric Smart
trc2.csv	1297941	2008-11-03 14:45:21	Tesla secures \$40 million in convertible financing
trc2.csv	1588233	2009-01-13 17:04:58	BRIEF-Tesla partnering with Daimler on electric Smart-Musk
trc2.csv	1588776	2009-01-13 19:50:41	AUTOSHOW-UPDATE 2-Tesla to supply battery for electric Smart
trc2.csv	1207751	2008-10-15 20:02:44	Tesla to delay new model, cut jobs to save cash
trc2.csv	1286239	2008-10-31 05:53:20	Tesla sees new financing, founder offers support
trc2.csv	1208014	2008-10-15 20:52:24	UPDATE 1-Tesla to delay new car, cut jobs to save cash

Navigation: 1 / 7

Selected record details:

msg_dt	message	story_text
2008-11-03 14:45:21	Tesla secures \$40 million in convertible financing	DETROIT, Nov 3 (Reuters) - Tesla Motors Inc said on Monday that its board had approved \$40 million in convertible debt financing to push ahead with stepped-up manufacturing of its battery-powered Roadster sports car. Last week, Tesla founder Elon Musk had said the electric car start-up was moving to close a new financing round from existing investors to bolster a cash balance that had dropped to \$9 million. San Carlos, California-based Tesla said it had commitments for the new convertible debt financing from most of its current, major investors. 'Forty million is significantly more than we need,' Musk said in a statement. 'However, the board, investors and I felt it was important to have significant cash reserves.' Tesla markets the \$109,000 Roadster, the first battery-powered performance car and a vehicle that has become a symbol of both the promise of electric cars and the financial risks of rushing the new technology to the market. Tesla has delivered

TREC 2. News content copyright Thompson Reuters.

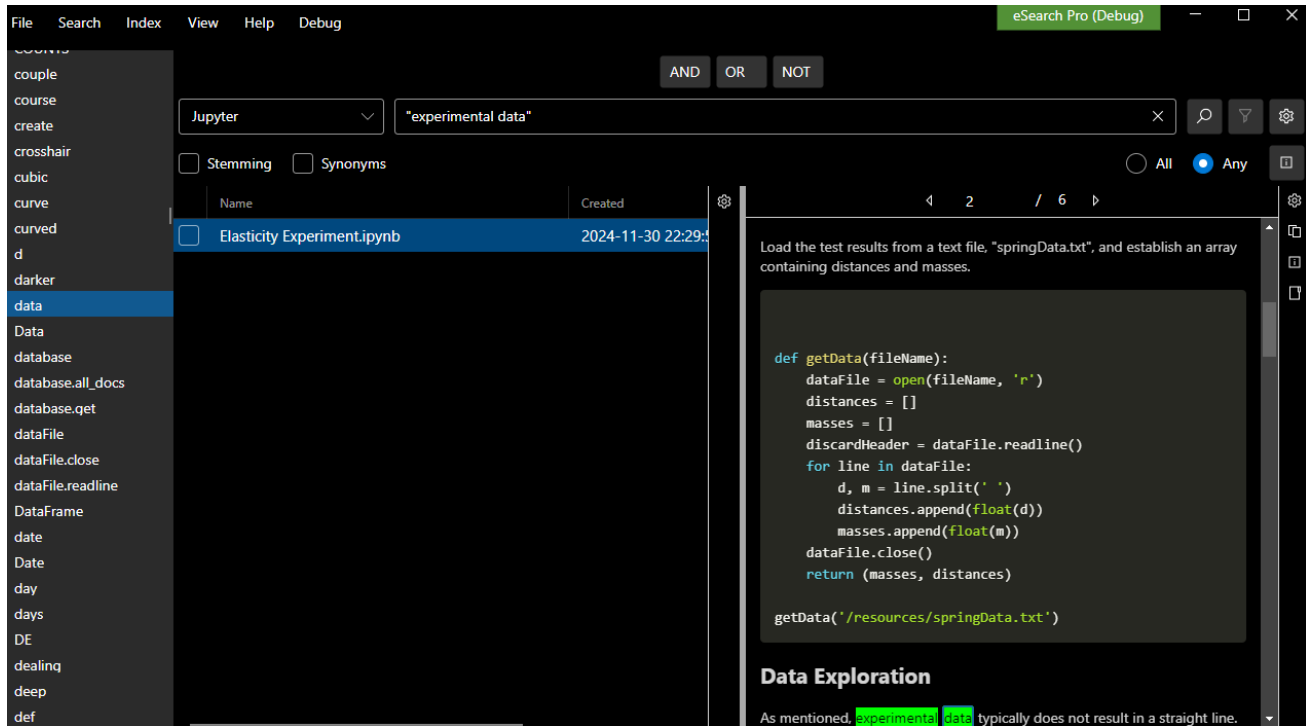
Test results:

eSearch Pro 1.0.9077. 8 minutes to index 1,800,370 records (2.67GB .csv file). Stop words = None. Case Sensitive = yes. No limit on word length or file size. Lenovo Think Station. Processor: 12th Gen. Intel Core i9.12900 2.4GHz. Windows 11 Pro 23H2 Storage 951GB, 680GB free. RAM 32GB (31.7 usable)

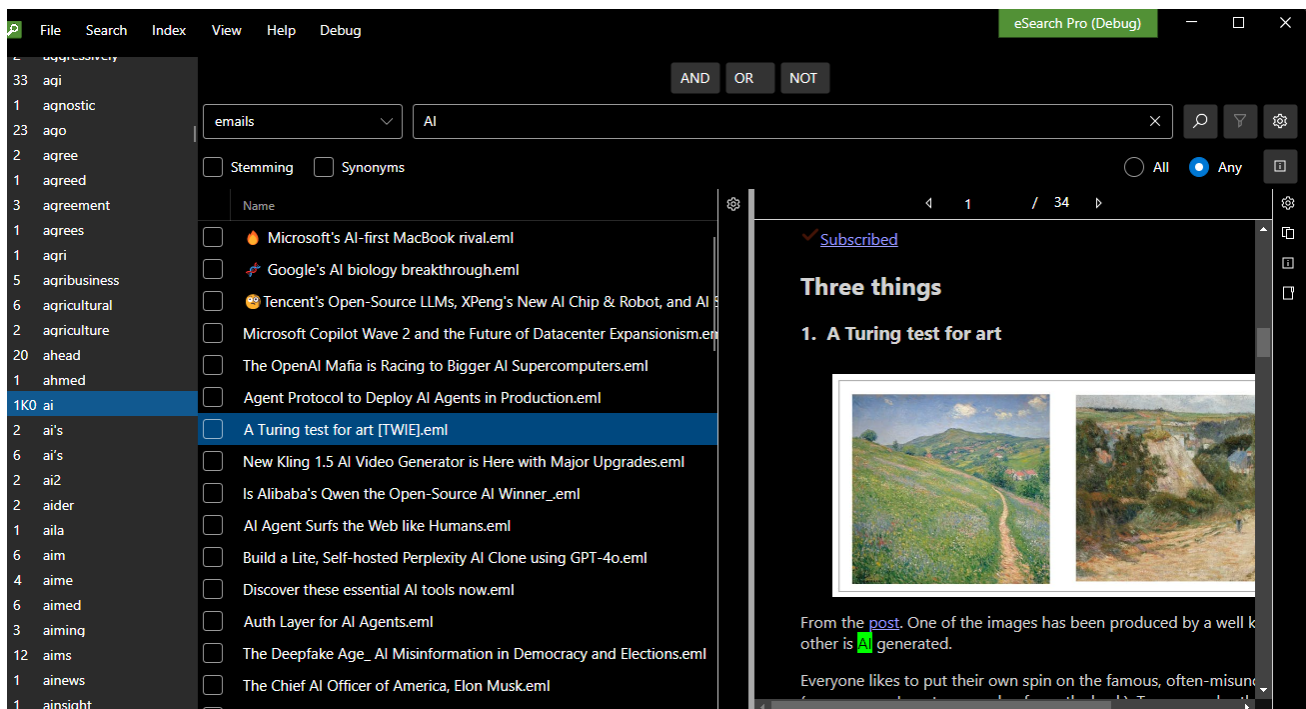
Research or Publish Newsletters and Notebooks.

There is an explosion of online material available through newsletters and Notebooks created by content creators on platforms like Substack, Bluesky, BeeHiiv, and others, not to mention the 20,000+ free papers published each month on arXiv.org and other pre-print services.

eSearch Pro now includes support for Jupyter Notebooks and Markdown files, as well as .eml files generated by Outlook's Save As feature.



Example Jupyter Notebook (.ipynb) file display showing hit highlighting.



Example .eml newsletter research. eSearch Portable - Search-only license lets you publish your own searchable Notebook collection.

30-day trial now available!

- You can download a fully functional trial version. After 30-days it will revert to Lite mode. eSearch Lite has some limitations on functionality, but you can continue to use it for free, for both private or commercial use, for as long as you wish.
- eSearch Pro is a work in progress, check back often for [new releases](#).
- eSearch Pro and Lite versions do not track usage or pass any data to us or any third party. There is error logging, which you can send to us if you need support.
- While in beta testing, both eSearch Pro and Lite are free of charge. You can download new trial versions of eSearch Pro at any time.
- If you have feedback, suggestions, questions, or need a different 'search' problem solved, please contact support@electronart.co.uk

Go to SearchCloudONE.com for free downloads & latest updates
on eSearch Pro!

ElectronArt Design Ltd.
124 City Road
London
EC1V 2NX