

E100 – SEARCHING FOR PEOPLE

This article describes how to search for people by name by using proximity operators, Soundex, synonyms or by using a “list of names”.

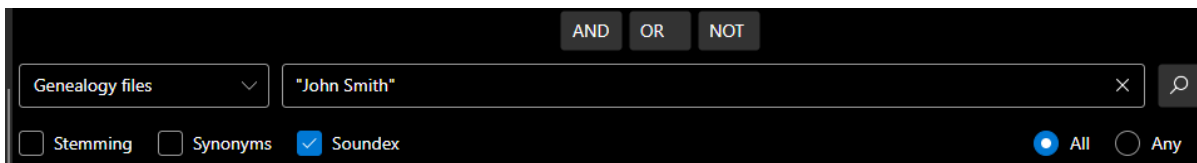
These are common processes used in genealogy, investigative journalism, anti-money laundering and law enforcement.

Contents

Introduction	3
Proximity operator	3
Soundex	4
Search Using a List of Names	5
Synonyms	6
Cross lingual	7
Political	8
Using AI online search	9
MCP Servers	9
Names in email format	10

Introduction

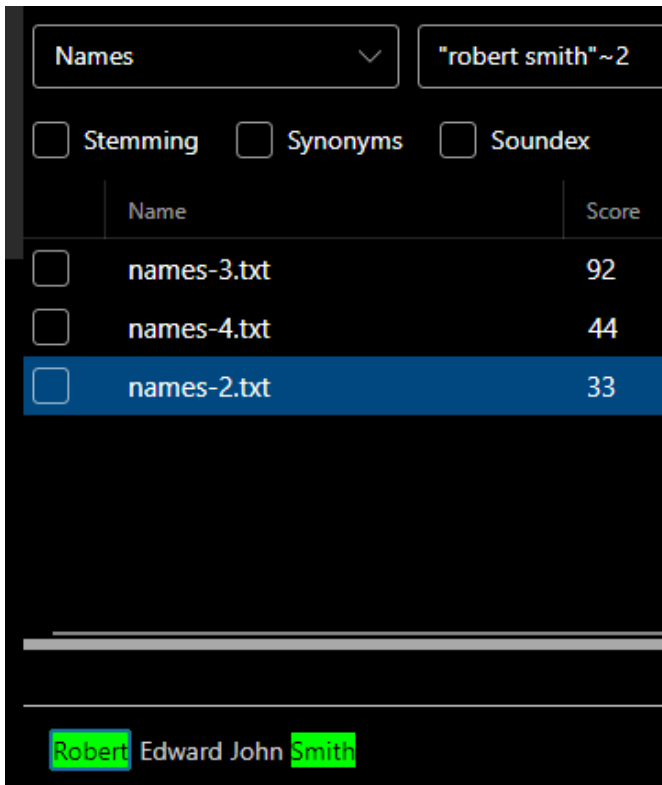
For name searching Stemming and Synonyms should not normally be selected.



Search interface showing a search for "John Smith". The interface includes a dropdown menu for "Genealogy files", a search input field with "John Smith", and buttons for "AND", "OR", and "NOT". Below the search bar, there are checkboxes for "Stemming", "Synonyms", and "Soundex". The "Soundex" checkbox is checked. At the bottom right, there are radio buttons for "All" and "Any".

Proximity operator

A useful technique when searching for names is to use the ~ proximity operator.



Search interface showing a search for "robert smith"~2. The interface includes a dropdown menu for "Names", a search input field with "robert smith"~2, and checkboxes for "Stemming", "Synonyms", and "Soundex". Below the search bar, there is a table with two columns: "Name" and "Score". The table lists three results: "names-3.txt" with a score of 92, "names-4.txt" with a score of 44, and "names-2.txt" with a score of 33. The "names-2.txt" row is highlighted. Below the table, there is a text area showing the results: "Robert Edward John Smith".

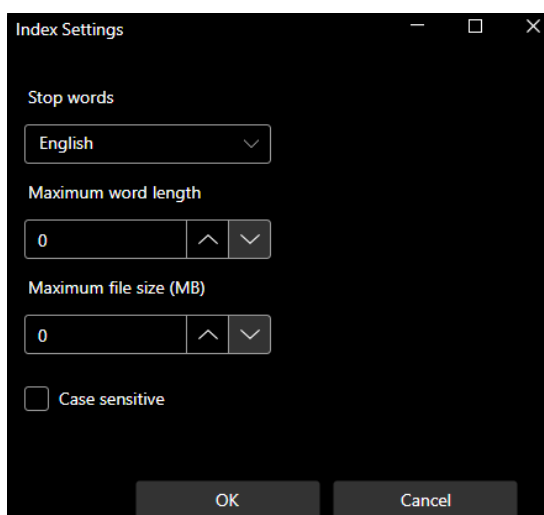
Name	Score
names-3.txt	92
names-4.txt	44
names-2.txt	33

For example, to ensure that a search for *Robert Smith* will also find *Smith, Robert* or *Robert Edward John Smith* you should search using

"Robert Smith"~2

Names-3: Robert Smith

Names-4: Smith, Robert



Index Settings dialog box. It contains the following settings:

- Stop words: English
- Maximum word length: 0
- Maximum file size (MB): 0
- Case sensitive: ☐

Buttons: OK, Cancel

The example shown above has the Case Sensitive option unselected in Index Settings. A search query for john smith will find John Smith or john smith.

Soundex

If you are unsure of the spelling of a name, use the Soundex option to expand the search to similar sounding names.

<https://en.wikipedia.org/wiki/Soundex>

A limitation of the Soundex algorithm is that words must start with the same letter to be considered a match. For instance, a search for “Smith” will also find “Smithe” and “Smythe”, and a search for “Philipp” will find “Phillipe” or “Philip”, but it will not find “Filip” because it begins with a different letter. If you have selected Stemming, it will also find “Phillips”, “Philips”, etc. which may or not be desirable depending on the purpose of your search.

Names

☐ Stemming ☐ Synonyms ☒ Soundex

	Name	Score
<input type="checkbox"/>	names -5.txt	337
<input type="checkbox"/>	names-3.txt	337

Rupert Smith

Example shows a search for “Robert Smith”, with Soundex option selected will also find “Rupert Smith”.

Names-3: Robert Smith

Search Using a List of Names

TARILIO simplifies your search process by allowing you to use a list of names in a text file, eliminating the need to type in names individually or create lengthy Boolean search terms like (thisName) OR (thatName) OR (otherName).

To use this feature:

1. Create a list: Enter one name per line in a Windows Notepad file.
2. Select Search option: In the Search Settings window, choose the "Use List" option.
3. Upload file: Browse to and select your name list file.
4. Click OK, the search query will appear in the search bar.
5. You can choose File > Export Search Results... to export all or selected results in a timestamped Excel tabbed .csv, .csv or .xml file format.

Search Settings

☒ Use List

D:\Person-Of-Interest.txt

Stemming Language

☐ English Porter

☒ English

Phonetic Analysis

None

Synonyms

☒ Synonym files:

Sample Names Cross Lingual

Synonyms

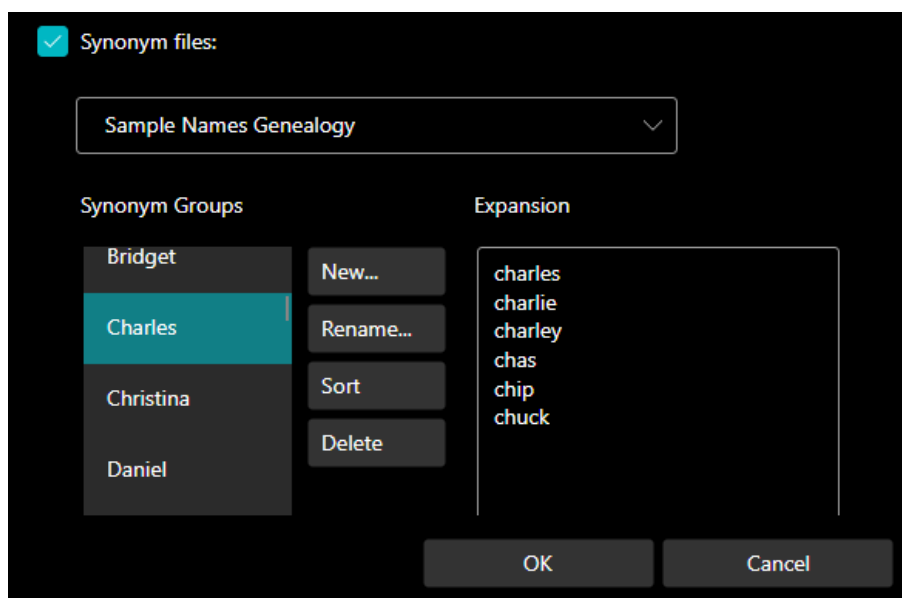
Where the files, email messages or website pages may contain the name(s) of interest in various formal or informal formats and may also be transcribed into other scripts such as Cyrillic or Arabic, it may be useful to list all the variants of the name(s) of interest into a Synonym file. TARILIO comes with several “name” sample files for different purposes.

Genealogy

The names_genealogy synonym file enhances TARILIO by broadening searches to include common diminutives of personal names.

Often when searching through census records, the names listed may not always match those on a person's birth certificate. For example, a common name on English Victorian census returns is Fanny, while the birth certificate might show Frances. Similarly, someone registered as Francis at birth might appear as Frank on the census.

The file contains common variants for many English names – in some cases the variants may include common variants from different English-speaking countries, for example Charles is most often Charlie or Chas in England, whereas in North America Chuck or Chip is common.



https://en.wikipedia.org/wiki/List_of_diminutives_by_language

https://en.wikipedia.org/wiki/Personal_name

<https://en.wikipedia.org/wiki/Hypocorism>

Cross lingual

The `names_cross_lingual` synonym file provides an example of how you can expand your search for alternative forms of Arabic names. This approach can also be used for names transliterated from Cyrillic to Latin script.

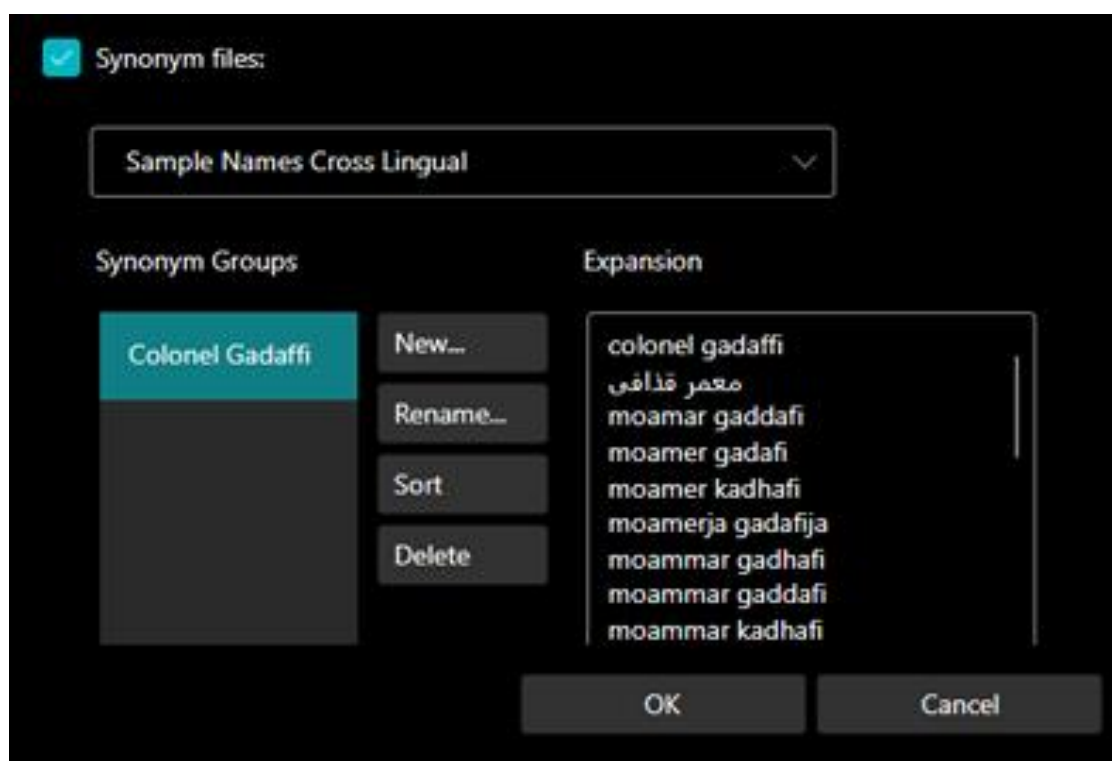
Transliteration involves changing each letter of a word from one script to another. this can be problematic when one language may not have an equivalent letter in its alphabet, in which case the letter with the closest pronunciation is chosen, which may lead to variant spellings in the same script in different languages.

Romanization involves transcribing Arabic to a Latin-based script based on the word's pronunciation. There are numerous standards (e.g., IPA, US Intelligence Community), which can result in variations between the target languages, as demonstrated in the sample file.

Refs:

en.wikipedia.org/wiki/Romanization_of_Arabic

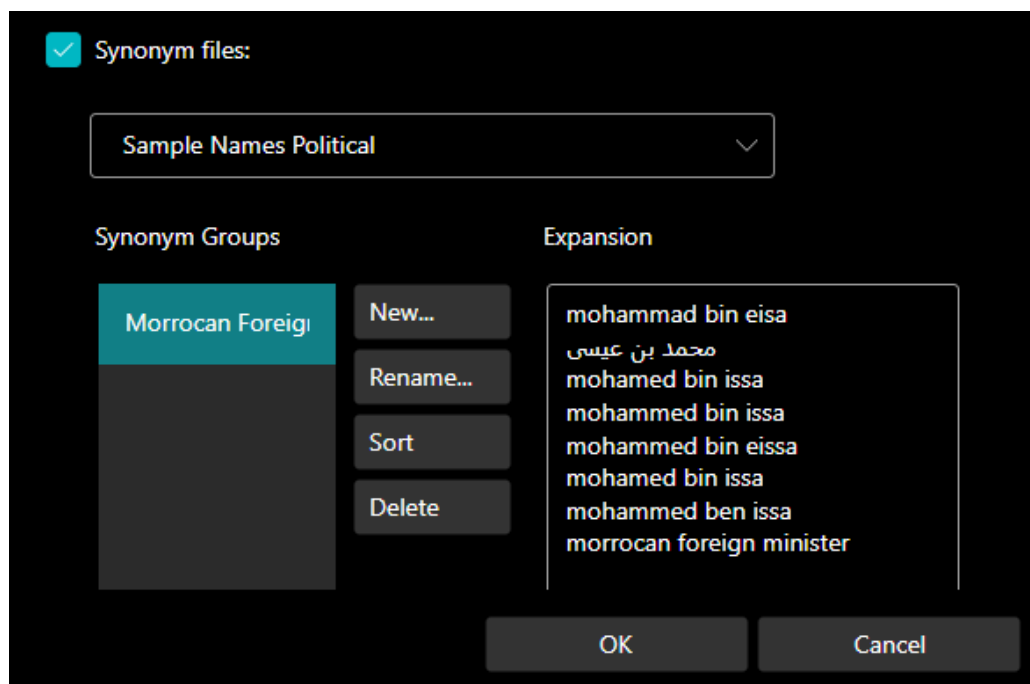
en.wikipedia.org/wiki/Transliteration



Political

The names_political synonym file demonstrates how you can expand your search in TARILIO by allowing searches based on a person's name (including possible aliases, transliterations, or Romanizations) or political office.

This type of search can be particularly useful when exploring news archives. However, it has the drawback that political offices are typically held for short durations, necessitating frequent updates to the file to ensure it includes the current office holder.



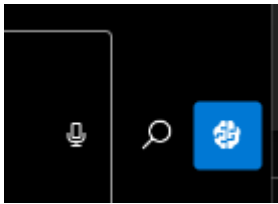
Refs:

<https://www.lawsociety.org.uk/support-services/advice/articles/peps/>

<https://transparint.com/blog/2016/08/01/7-open-sources-to-find-pep-information/>

<https://www.cia.gov/library/publications/world-leaders-1/>

Using AI online search



Switch to AI Mode by clicking the AI Mode button (blue button).

A screenshot of the 'Connect to API' dialog box. It has a dark background with white text. The 'Service' dropdown is set to 'Perplexity'. The 'API Key' field is empty with a dotted line indicating where to paste the key. The 'Model' dropdown is set to 'sonar-pro'. The 'Custom System Prompt' field contains the text 'Be precise and concise.' There is a 'Deactivate' button at the bottom right, and 'Test' and 'Cancel' buttons at the bottom.

From the AI menu, select AI connections. Select a Service and enter your API Key for the Model you wish to use*.

The System prompt can be edited if needed, but the default settings will generally provide satisfactory results.

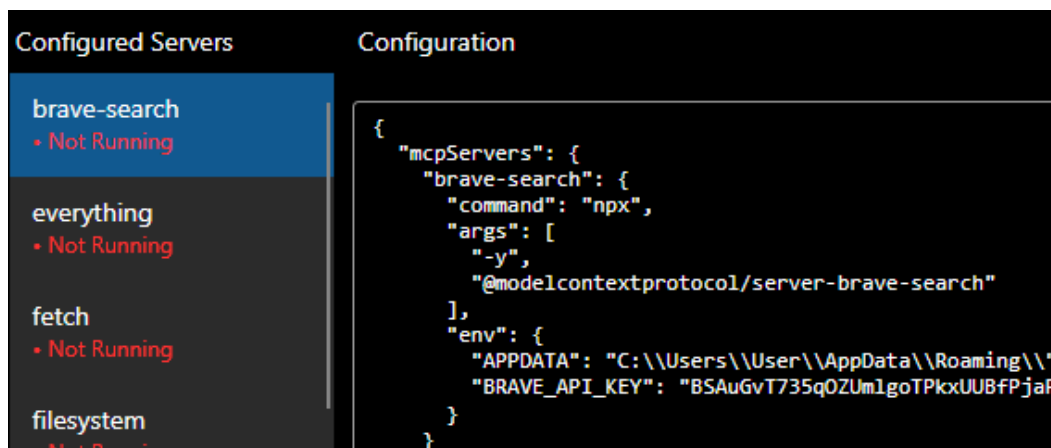
The Perplexity service combines up-to-date web search with large language models (LLMs), including reasoning models.

* Requires a Perplexity.ai API key. Pay as you go, no subscription required.

MCP Servers

Another way of searching online is to use an MCP server such as Brave-Search or Fetch. From the AI menu select MCP Servers.

Add an MCP Server by pressing the + button and pasting in the configuration



Names in email format

TARILIO splits email addresses on the @ character. If your files contain email addresses the word list will display the domain parts (contoso.com) separate from the email alias e.g. thomas.smith

Double-click on any domain or email alias in the word list to retrieve all contacts at the same organisation, or a specific email alias name at various domains.

To search on a complete email address enter it with quotes:

e.g. [thomas.smith@contoso.com](#)

oo0oo