dtSearch Desktop/Network
Indexing and Search techniques

T209 – SEARCHING FOR PEOPLE

dtSearch Desktop/Network is a powerful search tool used by professionals for a wide variety of tasks. This article describes how to search for people by name by using proximity operators, soundex, user thesaurus and by using a "list of words". These are common processes used in genealogy, investigative journalism, anti-money laundering and law enforcement.

Prerequisites
dtSearch Desktop/Network 7.88 or later
User Thesaurus Plus add-on product (optional).

Contents

Introduction

Before you attempt to carry out name searching using the methods in this article, we recommend that you follow the training course T207 to become familiar with setting up the initial indexing and default settings.

For name searching **Stemming** should normally be off, and the WordNet **Synonyms** and **Related words** should not be selected.

Proximity operator

A useful technique when searching for names is to use the **w/n** proximity connector; **w/n** is an abbreviation for *'within n words of'*.

For example, to ensure that a search for *Robert Smith* will also find *Smith, Robert* or *Robert Edward Smith* you should search using *Robert w/2 Smith*.

Sounds Like

If you are unsure of the spelling of a name, use the **Phonic** option to expand the search to similar sounding names; this is based on the well-known Soundex algorithm to find name spelling variants. You can select the **Phonic** checkbox to apply the Phonic option to all words or use the **#** symbol at the front of just those names you are unsure of.

A weakness of the Soundex algorithm is that words must start with the same letter to be considered a match. For example, a phonic search for #Smith will also find Smithe and Smythe, a search for #Philipp will find Phillipe or Philip but 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.

Search Using a List of Names

Many people are not aware that dtSearch Desktop allows you to search using a list of names in a text file, rather than having to type in names one at a time or having to write a long Boolean search term like (thisName) OR (thatName) OR (otherName), etc.

If you need to search for many names, simply enter them in a column in a Windows Notepad file or an Excel file; next from the **Search Menu** in dtSearch Desktop select "**Search for List of Words**" (**Ctrl+Shift+W**), choose the "**One word or phrase per line**" option then browse to your name list file.

This simple method may suffice where the name(s) in the files of interest are in a common format, but to ensure that variant formats can be found use the proximity operator **w/n** and choose the '**One Boolean expression per line**' option.

You can view the results directly or export the search results to an Excel file or plain text file for further investigation later.

User Thesaurus

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 **User Thesaurus** file.

The **User Thesaurus Plus** add-on product enables you to quickly create many files containing all alternative spellings, aliases, maiden names, diminutives, or nicknames and comes with several sample files.
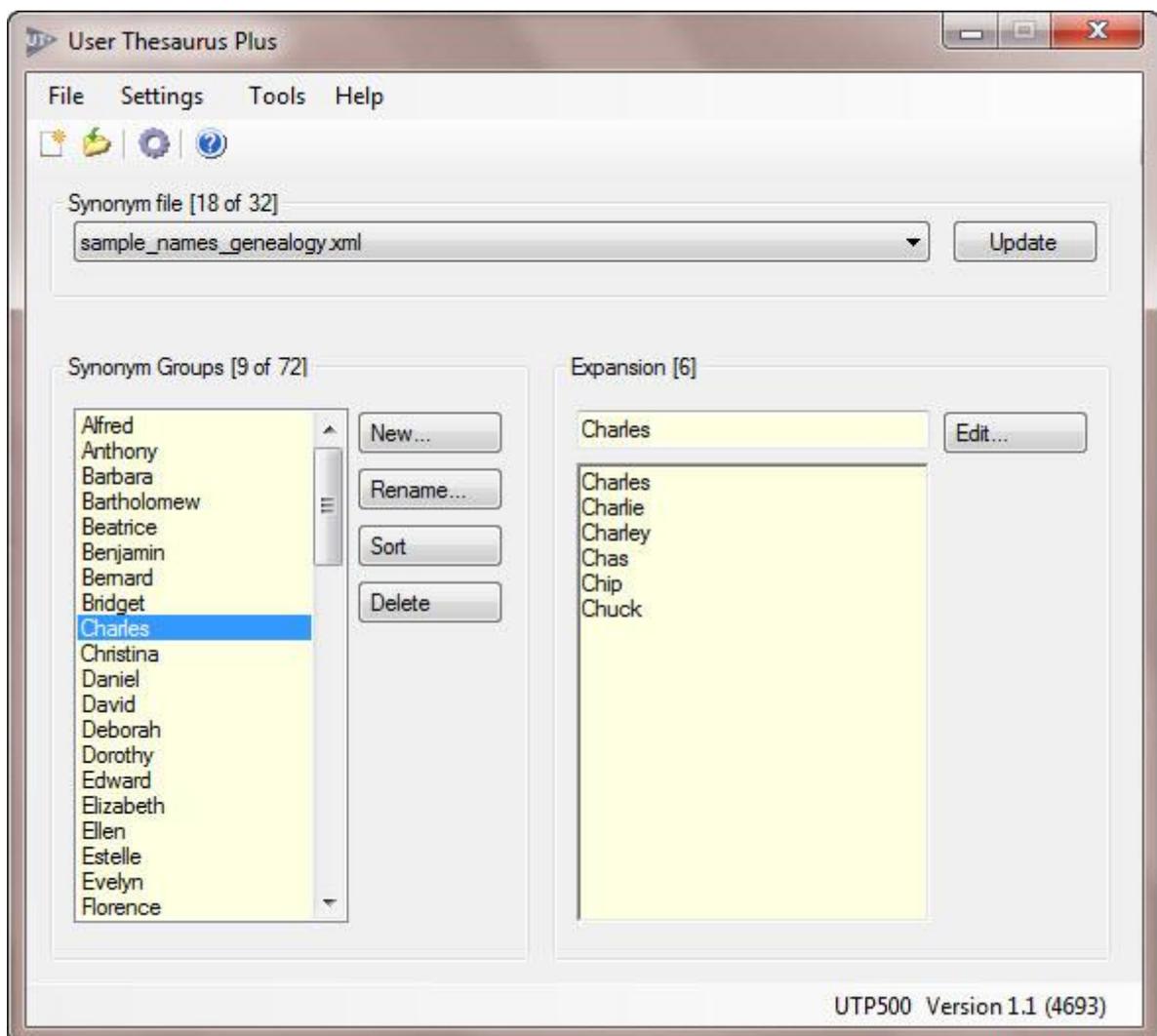
Genealogy

The **names_genealogy** thesaurus file will enable dtSearch to broaden your searches to include common diminutives of personal names.

Often when searching through census records the names that appear may not match the names that appear on the person's birth certificate, for example a common name appearing on English Victorian census returns is Fanny, whereas an inspection of the birth certificate shows they were named Frances at birth, similarly someone registered at birth with the name Francis would often appear as Frank on the English census.

The **names_genealogy** 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** thesaurus file gives an example of how you can expand your search for alternative forms of Arabic names in dtSearch. You can also do the same for names that have been transliterated say from Cyrillic to Latin script.
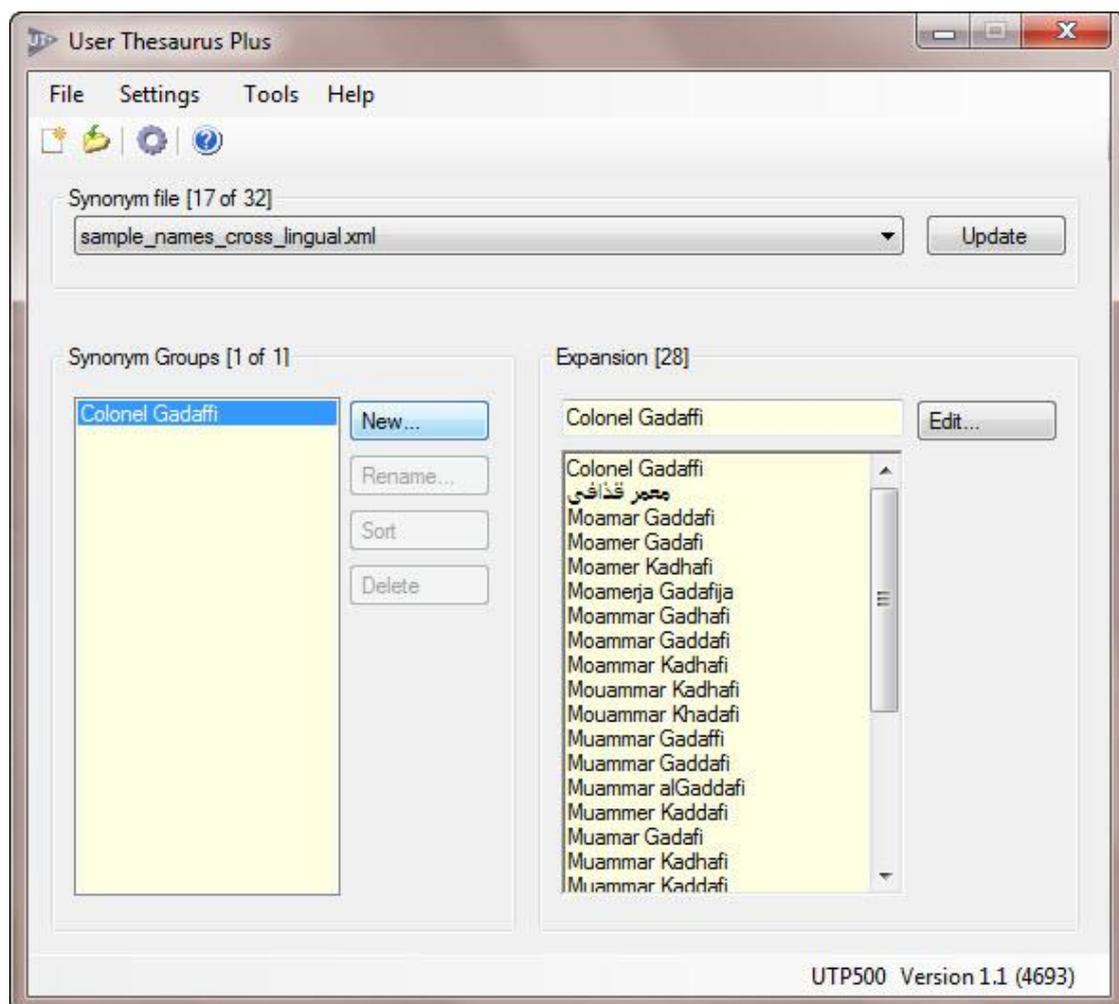
Transliteration is where each letter of a word is changed from one script to another, this can be problematic when one language may not have an equivalent letter in its alphabet, in which case the letter with the closest pronunciation is chosen which may lead to variant spellings in the same script in different languages.

Romanization is the transcription from Arabic to a Latin-based script based on the pronunciation of the word, there are many standards (e.g. IPA, US Intelligence Community) and this can lead to variants between the target languages as is evident in the sample file.
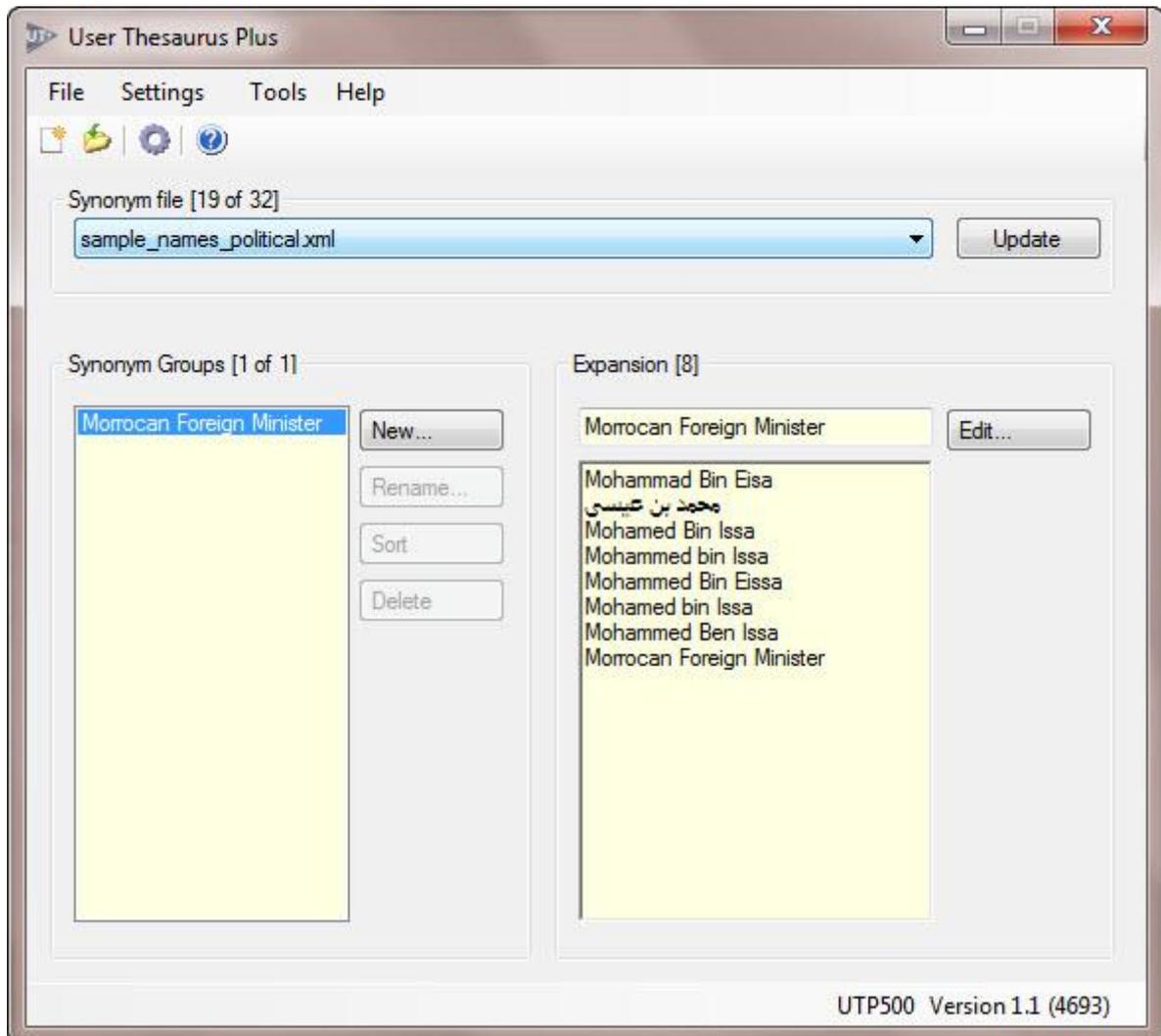
Refs:
en.wikipedia.org/wiki/Romanization_of_Arabic
en.wikipedia.org/wiki/Transliteration

The **names_political** thesaurus file gives an example of how you can expand your search in dtSearch to allow searching by a person's name (and possible aliases, transliterations or Romanizations) or political office.

This form of searching can be useful when searching news archives for example but has the disadvantage that the office is usually of short duration and therefor the file will need constant updating if you always want to search on news concerning the current office holder.

User Thesaurus Plus

File   Settings   Tools   Help

Synonym file [19 of 32]
sample_names_political.xml   Update

Synonym Groups [1 of 1]   Expansion [8]

Morrocan Foreign Minister   New...   Morrocan Foreign Minister   Edit...

                                 Rename...   Mohammad Bin Eisa
                                                محمد بن عيسى
                                 Sort   Mohamed Bin Issa
                                                Mohammed bin Issa
                                 Delete   Mohammed Bin Eissa
                                                Mohamed bin Issa
                                                Mohammed Ben Issa
                                                Morrocan Foreign Minister

UTP500   Version 1.1 (4693)

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/

https://www.dtsearch.co.uk/the-blog/blog/2014/april/30/searching-for-lists-of-names-offshore-leaks.aspx