



Indexing and limitations in PatBase & PatBase Express

This document highlights indexing and limitations when searching in PatBase.

Truncation

The following truncation is available in PatBase and PatBase Express:

Symbol	Function	Example	Limit
*	Unlimited Truncation	Right hand - Burn* finds burn, burnt, burning etc. Left hand - *inflammation finds antiinflammation, noninflammation etc.	Right hand – 100 in a query Left hand – 4 in a query
?	1 Character	Wom?n finds Woman and Women	2 in a word 100 in a query
%	0/1 Character and Fuzzy* Search	Colo%r finds Colour and Color	2 in a word 100 in a query
# or ~	Stemming	Apply# finds English grammatical variations, i.e. applying, applies etc.	100 in a query
!	0/1 Character		
_	Optional space		

When using truncation, please consider which words you are expecting to find when selecting the symbol to use. At times it is better to leave out truncation and simply enter the keywords you are interested in searching as this will limit noise.

** Fuzzy search is created when using %, this means that any letters before the % are treated as the stem of the word and PatBase/PatBase Express finds a maximum of one difference in the remainder of the word, this means a letter could be added, removed or changed. Please see example under searching Umlaut and Scandinavian letters below.*

Special Characters

When searching symbols, non-English characters etc. in PatBase, it is important to consider how they are indexed and what variations to search in order to find all relevant patent families.

Umlauts and Scandinavian letters

Umlauts and Scandinavian letters are now indexed as their long form and short form spelling variations:

Ä – AE or A
Ö – OE or O
Ü – UE or U
Æ – AE or A
Ø – OE or O
Å – AA or A

ß and Ñ

ß and Ñ are not indexed in PatBase, if these are found in the text they are indexed as follows:

ß - SS
Ñ – N

As the local resource might have already re-indexed the Umlaut or Scandinavian letter, it is important to search all possible variations.

For example, the company Mærsk, could be indexed as Maersk or Marsk. You could use the truncation symbol %, which looks for 0 or 1 character, but also creates a fuzzy search: e.g.:
PA=(ma%rsk)

This will find both Maersk and Marsk, however, as a letter could be removed, it would also find Mark. Hence, when searching Umlauts or Scandinavian letters it is best to write the two variations: e.g. PA=(maersk or marsk)

Non-Latin characters

It is possible to search Non-Latin characters by using the Non-Latin search form, or in PatBase, via the command line by preceding the search with [CC]:

e.g. [CN]: TAC=(生物精炼系统)

To search non-Latin assignees or inventors, use the search form by selecting non-Latin assignee/inventor from the drop-down menu by the assignee/inventor field.

Alternatively, use PANL or INVNL via the command line in PatBase.

Searching Numbers

It is important to add "" when searching numbers if they are separate, otherwise PatBase will see it as a search query number in your history, the correct way to search is:

TAC=(type "2" diabetes)

If a number is directly attached to another letter, it is not necessary to add "":

TAC=(type2 diabetes)

Greek Alphabet

When searching the Greek alphabet, simply spell out the letter:

A – Alpha

B – Beta

Γ – Gamma

Δ – Delta

Etc.

Stop Words

There are only 3 Stop Words in PatBase and PatBase Express: AND, OR and NOT.

If you want to search these as keywords instead of using them as Boolean operators, simply add "" : e.g. PA=(Johnson "and" Johnson)

It is not possible to highlight stop words.

Symbols

Search different symbols as follows:

Hyphens (-) – will search as if it is a space. E.g. hydro-chloric acid will search "hydro chloric acid"

Underscores (_) – will search words with a space and without a space e.g. (TAC=Non_Stick wf1 pan)

This will find the following three variations:

- Non stick pan
- Nonstick pan
- Non-stick pan

Apostrophes (') – same as hyphens e.g. Mozart's will search "Mozart s"

Circumflex (^) – remove from word

% - search as percent

+ and * can optionally be searched at the end of words e.g. C++ or A*