



Indexing and limitations in PatBase & PatBase Express

Truncation

The following truncation is available in PatBase and PatBase Express:

Symbol	Function	Example	Limit/Notes
*	Unlimited Truncation	Right hand - Burn* finds burn, burnt, burning etc. Left hand - *inflammation finds antiinflammation, noninflammation etc.	750 in a query
?	1 Character	Wom?n finds Woman and Women	
!	0/1 Character	Colo!r finds Colour or Color	
_	Optional delimiter	Pick_up finds Pick up, Pickup and Pick-up	10 in a word
%	Fuzzy* Search	Colo%r finds Colour and Color but also colon	2 in a word
# or ~	Stemming	Finds English grammatical variations, e.g. Apply~ will find applying, applies etc.	Only to be used at the end of a word. Works best on infinitive of a verb

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 keyword(s) 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 (character insertion/deletion/substitution/transposition) in the remainder of the word. Two % characters can be used to allow two edits. E.g. cond%uctor% will match conductors, conductr, conductor, conductor, conductors, conductors etc.

Please note: The optional delimiter () will not work in the assignee or inventor fields. This is due to the legacy of how the Browse index works. When using the Browse Index, the underscore operator is automatically added between assignee/inventor names in searches run from this tool. For example, if you are searching PA=(glaxo_smithkline) - This will search for families containing the exact assignee name Glaxo Smithkline and will not find variations such as Glaxo Smithkline Ltd.

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 to find all relevant patent families.

Non-English letters

Umlauts, Scandinavian, Eszett, eñe and Greek alphabet letters are indexed both in their original and as their long form, so either format can be used when searching in PatBase/PatBase Express.

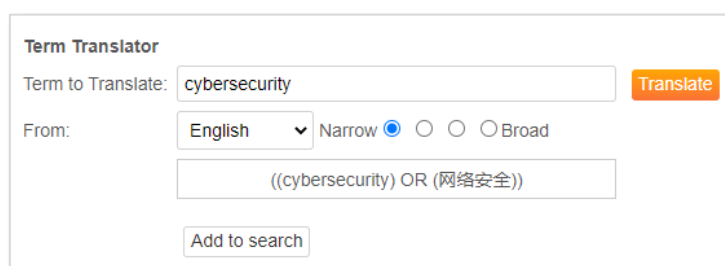
Umlauts	Long form
ä	ae
ö	oe
ü	ue
Scandinavian	Long form
æ	ae
ø	oe
å	aa
ß and ñ	Long form
ß	ss
ñ	n
Greek	Long form
α	Alpha
β	Bravo
Γ	Gamma
Δ	Delta
Etc.	

Non-Latin characters

Non-Latin and Latin characters are indexed together, it is therefore possible to combine different terms from different languages in a single search.

TAC=(人工智能 OR artificial intelligence)

Use the Term Translator on the Non-Latin search form to translate English terms to Non-Latin terms, e.g., Chinese or Japanese.



Term Translator

Term to Translate: cybersecurity Translate

From: English Narrow Broad

((cybersecurity) OR (网络安全))

Add to search

Non-Latin assignees or inventors can be searched along Latin assignee or inventors by using the search form or PA=/IN= in the command line.

PA=(三星 OR Samsung)

Traditional Chinese characters are indexed as simplified Chinese.

Searching Numbers

It is important to add quotation marks "" when searching standalone numbers, otherwise PatBase will interpret them as a search query number from your search history. To avoid this, either the number or the entire search term should be quoted:

TAC=(type "2" diabetes)

TAC=("type 2 diabetes")

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

TAC=(type2 diabetes)

Stop Words

PatBase and PatBase Express do not have any stop words i.e. non-indexed words. However, some words are used as operators in PatBase, and must be quoted to be

searched as keywords. These are: AND, OR, NEAR, NOT, NOTWn, NOTWFn, NOTWP, Wn, WFn, WP, nW (where n is a digit)

PA=(Johnson "and" Johnson)

Symbols

Search different symbols as follows:

The optional delimiter/underscore operator (_) – will search words with a space, punctuation and without a space. For example, searching for:

TAC=(Non_Stick pan)

This will find the following three variations:

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

Punctuation

Punctuation, such as #*+!:' ^ is not indexed and ignored in the search query. For example, Hyphens (-), Apostrophes ('), Circumflex (^) etc. when used within terms it will search as if it is a space. E.g. Searching:

hydro-chloric acid will search hydro chloric acid
Mozart's will search Mozart s
x^y will search x y

Please see exceptions below.

Punctuation	Long form	How to search
%	Percent	Either
&	and	Punctuation or quoted long form
< ≤ > ≥ =	Less than, greater than, etc	Quoted punctuation or long form
+ -	Plus or minus	Either
x	Multiply	Either
©	c	Either
®	r	Either
™	tm	Either
°	Degrees	Either

°C	Degrees centigrade	Either
°F	Degrees Fahrenheit	Either
X _n X ⁿ	Xn	Either

The following punctuation characters can be searched for when they occur at the end of a word if they are preceded in a query by a backslash #*+-%/:

C\+\+ will find C++

A* will find A*

nickel\-cadmium will find nickel-cadmium

(But not "nickel cadmium" or "nickel, cadmium")