

9 Searching in #LancsBox

Throughout the tool, #LancsBox offers powerful searches at different levels of corpus annotation using i) simple searches, ii) wildcard searches, iii) smart searches and iv) regex searches.

1. Simple searches are literal searches for a particular word (*new*) or phrase (*New York Times*). Simple searches are case insensitive; this means that *new*, *New*, *NEW*, *NeW* etc. will return the same set of results.
2. Wildcard searches are searches including one of three special characters *, <, > and =.

Special character	Meaning	Example of use
*	0 or more characters any word [with space]	<i>new*</i> [<i>new, news, newly, newspaper...</i>] <i>new *</i> [<i>new car, New York, new ideas...</i>]
>	larger than	
<	smaller than	
=	equals [combined with < and >]	

3. Smart searches are searches predefined in the tool to offer users easy access to complex searches; smart searches are unique to #LancsBox. These searches are used for searching for word classes (NOUNS, VERBS etc.), complex grammatical patterns (PASSIVES, SPLIT INFINITIVE etc.) and semantic categories (PLACE ADVERBS, HEDGES).
4. Regex searches are advanced searches that allow to search for any combination of characters. Any expression enclosed in forward slashes (/) is interpreted as regular expression. #LancsBox supports pearl-compatible regular expressions.

Regex	Explanation	Regex	Explanation
Word	A string of characters (case sensitive)	a{3}	Exactly 3 of a
/word/i	A string of characters (case insensitive)	a{3,}	3 or more of a
/word\./p	Punctuation search: A string of characters followed by full stop (case sensitive)	a{3,6}	Between 3 and 6 of a
[abc]	A single character either a, b or c.	\d	Any digit
[^abc]	Any single character except: a, b, or c	\D	Any non-digit
[a-z]	Any single character in the range a-z	\w	Any word character (letter, number, underscore)
[a-zA-Z]	Any single character in the range a-z or A-Z	\W	Any non-word character
[0-9]	A single number in the range 0-9		
.	Any single character		
(a b)	a or b		
a?	Zero or one of a		
a*	Zero or more of a		
a+	One or more of a		