# 10  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, iv) regex searches and v) batch 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.
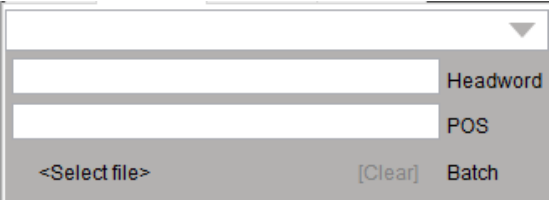2.  Wildcard searches are searches including one of three special characters *, <, > and =.

| Special character | Meaning | Example of use |
|---|---|---|
| * | 0 or more characters | new* [*new, news, newly, newspaper*…] |
|  | any word [with space] | new *[*new car, New York, new ideas*…] |
| > | 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 perl-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 |  |  |

5.  Batch searches allow to search for multiple search terms recursively and saving the results automatically; #LancsBox supports both simple and complex batch searches. Batch searches can be used in KWIC, GraphColl and Whelk modules when the corpora are tagged. Here is how batch searches work.

a)  Click on the down arrow in the search box to activate Advanced search options. The last option is a batch search. Click on 'Batch'.



| | Headword |
| --- | --- |
| | POS |
| <Select file> | [Clear]  Batch |

b)  Navigate to and load a text file with the appropriate search terms, one per line. Simple search terms include a list of word forms to be searched; complex search terms are defined via a combination of criteria such as word form, pos tag, headword etc… Consecutive criteria need to be present on the same line separated by tab (\t) in the following order: label – wordform – headword – pos – user tag. This is best achieved by creating the file with advanced batch search terms in Excel or Calc. Examples of simple and complex searches can be seen below.

| Simple batch search: each search term on a separate line | Complex batch search: label – wordform – headword – pos – user tag (tab separated) | | | | |
| --- | --- | --- | --- | --- | --- |
| my | | A | B | C | D | E |
| cat | 1 | multiple-simple | /cat\|dog\|mouse/ | | | |
| go | 2 | headword | | go | | |
| went | 3 | pos | | | N* | |
| | 4 | user | | | | Mytag |
| | 5 | combibnation | /going\|went/ | go | /V.*/ | Mytag |

c)  Once the file with search terms is loaded, click on the 'Search' button ( Search ) and navigate to the location where the results will be saved.