Index Search Overview

FARSIGHT

General

Text Processing

All searchable words must consist of alpha-numeric characters. All characters outside of this definition are treated as word separators. The following list includes examples and is not exhaustive:

- Non-printable characters e.g. `blanks`, `tabs`, `new lines`
- Printable characters for punctuation e.g. `.?!,:;-_[]{}()` ´'"`
- Currency symbols e.g. `$£€`
- Other printable special characters e.g. `+*<>&/\^~`

ⓘ Farsight does not filter any common words typically referred to as "noise"- or "stop words" as seen in other products. This can help to retain precision when doing searches for phrases as removing common words like `in`, `as`, `we` can lead to ambiguous results or false positives.

Case Sensitivity

Searches are executed case insensitive and a search for `energy` will find all variations e.g. `Energy`, `energy`, `ENERGY`, `EnErGy`.

Accented Characters

English uses diacritics sparingly compared to many other languages, particularly those derived from Latin. For compatibility all diacritics are removed to map any accented character to their base character. A search for `énergie` is the same as `energie` and will hit on both notations.

Term handling

Single term

Enter a single word e.g. `energy` to show documents containing that word.

Phrase term

Enter multiple words e.g. `energy plan` to show documents containing that exact phrase. For better readability it is recommended to enclose phrases in straight quotes e.g. `"energy plan"`.

Multiple terms

Terms and phrases can be linked with logical operators which is explained in the next section.

Operators

ⓘ The operators can be searched when enclosed by quotes. Searching for example `"oil and gas"` will find that exact phrase. Furthermore operators are not case sensitive but are are noted in capital letters here for better readability.

AND

All connected terms are required to be present. For example `energy AND plan AND resources` requires to have all three terms to match.

OR

Any single or multiple occurrence would yield a match e.g. `energy OR plan OR resources`.

NOT

Negation of the immediately following term or phrase.

- `NOT plan` will yield documents not having the word plan present.
- `energy AND NOT plan` will yield documents having the term energy present and term plan absent.
- `energy AND NOT "energy plan"` will show documents with term `energy` but any document containing phrase `"energy plan"` will be excluded.

⚠ Please note that exclusions introduced by AND NOT are handled strictly. Based on the last example a document containing `"energy plan"` will remain excluded even if the document contains `energy` somewhere else and not in conjunction with `plan`.

W/n

Find a term within distance (n) of another term with no specific order. For example `energy W/5 plan*` is the same search as `plan* W/5 energy` and will yield documents containing the following sentences:

- ... energy plan ...
- ... planned energy ...
- ... energy was distributed according to plan ...
- ... plants deliver insufficient energy ...
- ... energy distribution requires extensive planning ...

ⓘ In Farsight proximity searches with phrases work slightly different than in other products. `"energy plan" W/5 oil` will find the words `energy`, `plan` and `oil` if they are close together (a phrase can be re-arranged with `5` shifts to contain all three terms). This means that the terms do not strictly have to appear in a given order and natural language variations are captured too.

Index Search Overview

FARSIGHT

Wildcards

The samples below demonstrate the use of wildcards at the end of the term. In fact, wildcards are permissible at any position including the beginning of the term.

?

Use question mark to substitute one indistinct character. For example `plan?` will hit on plans, plane, plant but not plan or planned.

*

Use asterisk to match optional indistinct characters. For example `plan*` will hit on plan, plans, plant, plane, plants, planes, planned.

=

Use the equal sign to match any single digit. For example `20==` will hit on 2000, 2001, 2002, [...], 2098, 2099.

Fuzziness

%n

Fuzzy search allows matching terms that are within a distance of n from the search term. The distance is limited to be either 1 or 2. A distance of 1 means the match can differ by a single character insertion, deletion, substitution, or transposition. For example, the query `plan%1` might return results like pan, plain, plank, or plant.

The effectiveness of fuzzy search increases with longer words. For instance, searching for `energy%1` may yield few results beyond exact matches or simple typos. However, increasing the distance with `energy%2` may retrieve matches like synergy, emerge, entry, or every. While fuzziness can help catch typos or near matches, it may also significantly reduce precision by including unrelated terms.

Grouping

( )

Use parenthesis to clarify processing order in complex queries e.g. `energy AND (plan OR resources)`.

Regular Expressions

ⓘ While general regular expression syntax is supported, only word-level patterns are applicable. Multi-word and multi-line expressions are not supported due to the word-centric nature of the search index. Additionally, all indexed text is lowercased, so expressions relying on uppercase character classes will not match any results.

General

A regular expression term is started with straight quotes and two pound signs to contain the pattern and is closed by straight quotes: `"##..."`.

Given the complexity of regular expressions, only a selected subset of commonly used patterns is shown below.

Character Classes

Shorthand

`.`	any character
`\d`	digit
`\D`	not digit

Unicode

`\pX`	unicode character class identified by abbreviation
`\p{Greek}`	unicode character class identified by name
`\PX`	negated unicode character class identified by abbreviation
`\P{Greek}`	negated unicode character class identified by name

Bracket

`[0-9]`	digit character class
`[a-z]`	matching any character in range a-z
`[abc]`	matching either a, b or c
`[^abc]`	matching any character except a, b and c

POSIX

`[[:digit:]]`	digit character class
`[[:alpha:]]`	alphabet character class
`[[:^alpha:]]`	negated alphabet character class

Repetitions

`a?`	zero or one of a
`a*`	zero or more of a
`a+`	one or more of a
`a{n}`	exactly n a
`a{n,m}`	at least n a and at most m a
`a{n,}`	at least n a