Search Syntax

PhiloLogic4's query syntax has 5 basic operators:

- the plain token, essentially, any word at all, split on space, e.g. `usura`
- the quoted token--a string in double quotes, which may contain a space, e.g. `"usurarum voraginem"`
- the range--two tokens separated by a dash, e.g. `1350-1450`
- boolean OR, represented grep-style as `|`, e.g. `usura | vsura`
- boolean NOT, represented SQL-style as `NOT`, e.g. `usur.* NOT usurp.*`

This syntax is the same, but interpreted slightly differently, for the two different types of text query fields: word search and metadata search.

Word Searches

Full-text word search is unique in having the concept of a "term", which is either a single plain/quoted term, or a group of plain/quoted terms joined by `|`, optionally followed by `NOT` and another term-like filter expression:

- plain terms are evaluated without regard to accent. Regexes are permitted.
- quoted terms are case and accent sensitive. Regexes are permitted.
- the range is not operational. In the future, stub this out to make hyphenated search terms less of a pain to escape.
- `OR` can conjoin plain and quoted tokens, and precedes evaluation of phrase distance.
- `NOT` is a filter on a preceding term, but cannot stand alone: `u.* NOT usura` is legal, `NOT u.*` is illegal

Metadata Searches

Metadata search does not support phrases, but supports more sophisticated Boolean searching:

- plain tokens separated by spaces have an implied AND between them, but are treated as position-independent tokens.
  Regexes are permitted, but will not span over the bounds of a token.
- quoted tokens must now match against the ENTIRE metadata string value in the database, including spaces and punctuations.
  It will not match a single term within a larger string, no matter how precise. Regexes are permitted
- range allows for numeric and string ranges on all metadata fields.
- `OR` can still be used to conjoin plain tokens, preceding the implied Boolean AND, as well as quoted tokens.
- `NOT` is still available as both a filter, or a stand-alone negation: `diocesan NOT compilation` is legal, so is `NOT diocesan`

Regexp syntax

Basic regexp syntax, adapted from the egrep regular expression syntax:

- The character `.` matches any single character except newline.
- Bracket expressions can match sets or ranges of characters: `[aeiou]` or `[a-z]`, but will only match a single character unless followed by one of the quantifiers below.
- `*` indicates that the regular expression should match zero or more occurrences of the previous character or bracketed group.
- `+` indicates that the regular expression should match one or more occurrences of the previous character or bracketed group.
- `?` indicates that the regular expression should match zero or one occurrence of the previous character or bracketed group.

Thus, `.*` is an approximate "match anything" wildcard operator, rather than the more traditional (but less precise) `*` in many other search engines.