Symbols

() match but will not include in the match, not displayed

?<= Positive lookbehind ("assert that what precedes is...")

?<! Negative lookbehind ("assert that what precedes is NOT...")

?= Positive lookahead ("assert that what follows is...")

?! Negative lookahead ("assert that what follows is NOT...")

. a word or a punctuation mark, exactly one

[^。] one character, but not a 。 preferred if you aim at shorter phrase within a sentence.

+ equals to {1,} i.e. once or more time

* equals to {0,} i.e. zero or more time, greedy

? equals to {0,1} i.e. zero or one time, non greedy

.* equals to one character and then as many as possible, greedy.

.*? equals to one character and then as few as possible, non greedy.

[一-龥]+ matches one or more Chinese characters

[一-龥]{2} matches exactly two Chinese characters

[一-龥]{2,4} matches between 2 and 4 Chinese characters

(?=.*鄉約)(?=.*教案) matches these two terms in any order - very slow searching!

Blocks

[一-龥] Chinese characters block

[[，。、；：？！…—―·「」『』（）〔〕【】《》〈〉]] for chinese punctuation marks block

Chinese punctuation mark

[\x{3000}-\x{303F}] = match any Chinese punctuation mark

[\x{3000}-\x{303F}]夫 = any chinese punctuation mark and a 夫 follows it

(?<![\x{3000}-\x{303F}])夫 = match '夫' where the preceding character is not a Chi punctuation mark

Phrases

(?<=。)(.*?禁.*?戲)(?=。) = from a phrase contains 禁, not necessarily at the beginning of sentence, till a phrase contains 戲, both fullstops not included in display

地[保|方] = match 地保 or 地方

地(?=保|甲) = match 地保 or 地甲 but just highlight 地

[一-龥] 夫 = match any chinese character and a 夫 follows it

(?<![一-龥])夫 = match '夫' where the preceding character is not a Chinese character

夫(?![一-龥]) = match 夫 not followed by a chinese character

(?<=[^一-龥])[一-龥]夫(?![一-龥]) = will display two chinese characters. a non chinese character (e.g. punctuation mark) which will not be displayed, follwed by a chinese character, a 夫, not followed by any chinese character which will not be displayed.

[^。] one character, but not a 。 preferred if you aim at shorter phrase within a sentence.

[^。]{2} two characters, but both not 。

. a word or a punctuation mark, exactly one

\n equals to new line

以[^。]*?而.*?。 以...而...。 without punctuation between 以 and 而

以([^。]*?)而(.*?)(?=。) capture 以...而... ends with but does not display。

[^。，；：！？「\n]{0,}者   capture from the first character all the way to the first 者

\+ = search for plus sign instead of treating + as a variable

(人.){2,} Highlight all phrase with 人 X 人 X while the pattern of 人 X repeats twice or more times.

(人.){2,4} 人 X repeats two to four times

(人.){2} 人 X repeats exactly two times

去後.*?據.*?。 starting with "去後", followed by any number of words until the first occurrence of "據", and then continues to the first "。"

去後.*據.*。 [not recommended] starting with "去後", followed by any number of words until the last occurrence of "據", and then continues to the last "。"

等[^。]到[^。] Highlight a 4 characters phrase with 等 X 到 X while excluding 。

[^。，；：！？「\n]{1,}者，[^。；？]{1,}也[，；。！？] Highlight "...者...也" pattern. captures from beginning of a Chinese character all the way to the first 者, and all the way to the first 也 including the ending punctuation, except there is a 。；? that breaks it.

主教.{0,10}司鐸 highlight any phrase from 0 to 10 characters between 主教 司鐸 where dot means a single character, {0,10} means the number of that character can be from 0 to 10

(?<=。|^)(?=[^。]*主教)(?=[^。]*司鐸)[^。]*。 包含在兩個句號之間有 "主教" 和 "司鐸"（任何順序）這兩個詞。

(?<=。|^)(?=[^。]*哭|稟|訴)(?=[^。]*前|來|堂)[^。]*。 包含在兩個句號之間有 "哭或稟

或訴" 和 "前或|來|堂" （任何順序）這兩個詞。

以.{2,4}而.{2,4} = 以 followed by 2-4 words

[稟訴][堂前來].|[堂前來].[稟訴] Match Any four characters phrase with "稟 or 訴" in second/fourth place and "前來" or "X 堂" as first/third place Example: "哭稟來堂", "來堂哭稟"

(?<=^|[\s.,;!?，。；！？])(?=[^\s.,;!?，。；！？]*[哭稟訴])(?=[^\s.,;!?，。；！？]*[前來到堂])[^\s.,;!?，。；！？]+(?=[\s.,;!?，。；！？]|$)

Description: Matches a phrase at least one character from [哭稟訴] and at least one character from [前來到堂] Example: "哭稟來堂"

(?=.*主教)(?=.*司鐸)[^\s.,;!?，。；！？]+

Description: Ensures that both "主教" and "司鐸" appear within the same uninterrupted sequence of characters.

^(?!.*主教).*(司鐸|總鐸)

Description: Matches any string containing "司鐸" or "總鐸" as long as "主教" does not appear anywhere in the string.

(^|[\s.,;!?，。；！？])([^主教\s.,;!?，。；！？]*司鐸[^主教\s.,;!?，。；！？]*總鐸[^主教\s.,;!?，。；！？]*|[^主教\s.,;!?，。；！？]*總鐸[^主教\s.,;!?，。；！？]*司鐸[^主教\s.,;!?，。；！？]*)(?=[\s.,;!?，。；！？]|$)

Description: Matches any string containing "司鐸" AND "總鐸" as long as "主教" does not appear anywhere in the phrase

(?<=^|[\s.,;!?，。；！？])(?=[^\s.,;!?，。；！？]*主教)(?=[^\s.,;!?，。；！？]*(司鐸|總鐸))(?![^\s.,;!?，。；！？]*教民)[^\s.,;!?，。；！？]+(?=[\s.,;!?，。；！？]|$)

Description: Ensures that the unit contains "主教" and either "司鐸" or "總鐸," but excludes any unit containing "教民." Example: "主教總鐸"