<1>

<5>

<4>

<aa>

<-1>

how to identify new page?:

- regex
- check first char of each line and last char :: " " "
  check for < && >
  check if middle is a digit

add spaces before and after each word

strip all punctuation & white space

<1> hello my name is jonathan

Key word
- jonathan

<1> ␣hello␣ my␣ name␣ is␣ jonathan␣

check if string contains keyword w/ space

␣jonathan␣