①변경된 점들

1. 정규식 2차 수정

②추가사항

1. 정규식 2차 수정(노란색 부분이 변경된 것들)

사진에 경우 url, 가로길이, 세로길이 파싱하여 저장한다.

pics += re.findall('<img src="(.\*?)".\*?data-width="(.\*?)" data-height="(.\*?)".\*?class="se-image-resource" />', html)

스티커는 그대로 url

sticker += re.findall('<img.\*? src="(.\*?)".\*?class="se-sticker-image" />',html)

텍스트는 그대로 텍스트 그 자체

text += re.findall('<span.\*?>(.\*?)</span>',html)

쓸대없는 문자열들이 많이 빠졌지만 그래도 몇몇은 남아있음