Docx 파일 복구 연구

1. 덤프 파일 03_dump_txt.img 에서 docx 해더 시그니처(504b)를 찾아봄

```
by@ubuntu:~/dump$ sigfind -b 4096 504b 03_dump_txt.img
Block size: 4096  Offset: 0  Signature: 504B
Block: 33192 (-)
Block: 35328 (+2136)
error reading bytes 507619
```

⇨ 다음과 같이 결과가 나오는 것을 알수 있다.

2. 덤프 파일 03_dump_txt.img 에서 docx 끝에 존재하는 시그니처 504B05 를 찾아봄

```
by@ubuntu:~/dump$ sigfind -b 4096 504b05 03_dump_txt.img
Block size: 4096  Offset: 0  Signature: 504B05
error reading bytes 507619
```

⇨ 결과 값이 나오지 않는다.

⇨ 혹시 존재하지 않는 걸까 해서 docx 파일의 끝부분을 확인해봄

```
003800f0: 2e78 6d6c 504b 0506 0000 0000 2000 2000  .xmlPK...... . .
00380100: 4908 0000 abf8 3700 0000 0000 0000 0000  I.....7.........
00380110: 0000 0000 0000 0000 0000 0000 0000 0000  ................
```

⇨ 다음과 같이 존재하는 것을 확인

⇨ 결론은 sigfind 로는 끝을 확인할 수 없다.

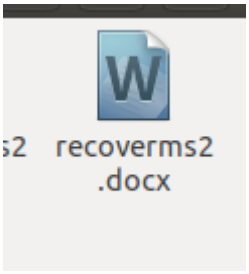
3. 0 padding 이 포함 된 이상 dump 파일 안에 word(docx) 의 자료가 들어가도 복구 불가능하다.

```
00380fd0: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00380fe0: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00380ff0: 0000 0000 0000 0000 0000 0000 0000 0000  ................
by@ubuntu:~/dump$ dd if=03_dump_txt.img bs=4096 count=897 skip=35328 >recoverms2
```

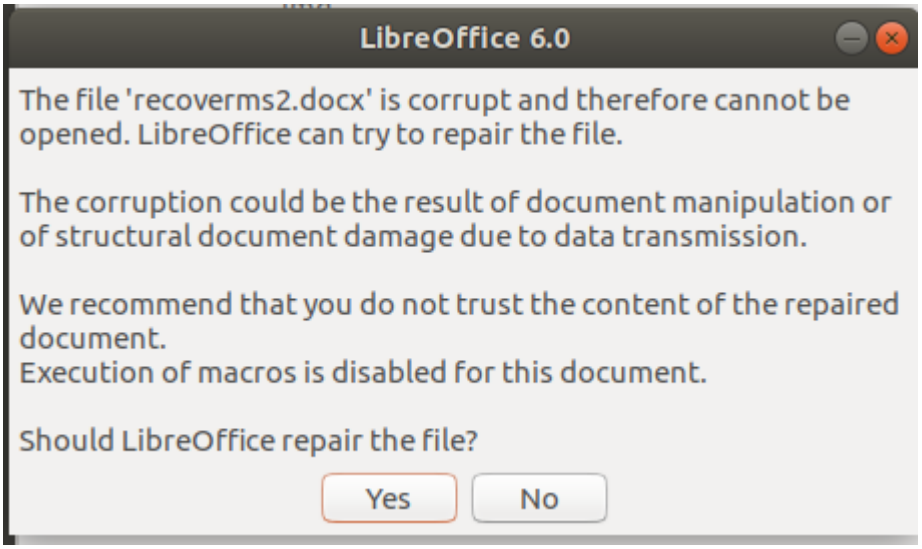
⇨ 다음과 같은 명령어로 0 패딩이 포함한 word 덤프 파일을 만듬.

⇨ Cat 으로 복구 시도.

```
by@ubuntu:~/dump$ cat recoverms2 > recoverms2.docx
```



⇨ 하지만 열려고 하는 순간 다음과 같은 경고문이 뜬다.



⇨ 결론 0 padding 을 잘라서 완벽한 순수 word 파일을 만들지 않은 이상 복구 불가능하다.

4. Docx 구조.

```
by@ubuntu:~/dump$ xxd recoverms2
00000000: 504b 0304 1400 0808 0800 f604 484f 0000  PK..........HO..
00000010: 0000 0000 0000 0000 0000 1000 0000 776f  ..............wo
00000020: 7264 2f66 6f6f 7465 7231 2e78 6d6c a595  rd/footer1.xml..
00000030: db6e 9c30 1086 9fa0 ef80 7cbf 0b44 490f  .n.0......|..DI.
00000040: 68d9 5c74 d5a8 522f 5669 fb00 1363 c08a  h.\t..R/Vi...c..
00000050: 4f1a 1be8 be7d cd39 bb1b 4584 7081 e519  O....}.9..E.p...
00000060: cff7 8fc7 83d9 ddff 9322 a819 5aae 554a  ........."..Z.UJ
00000070: e26d 4402 a6a8 ceb8 2a52 f2f7 cf8f cd57  .mD.....*R.....W
00000080: 1258 072a 03a1 154b c989 5972 bfff b46b  .X.*...K..Yr...k
00000090: 92dc 61e0 8395 4d24 4d49 e99c 49c2 d0d2  ..a...M$MI..I...
```

Offset	Decription
0 – 3	local file header signature (0x04034B50)
4 – 5	version needed to extract (0x1400)
6 – 7	general purpose bit flag (0x0808)
8 – 9	compression method (0x0800)
10 – 11	last modify file time (0xf604)
12 – 13	last modify file date(0x484f)
14 – 17	crc-32 (0000 0000)
18 – 21	compressed size (0000 0000)
22 – 25	uncompressed size (0000 0000)
26 – 27	file name length (0x1000)
28 – 29	extra field length (0000)
variable size	file name ● 776f 7264 2f66 6f6f 7465 7231 2e78 6d6c a595 ● word/footer1.xml..
variable size	extra field