



한글 문서 파일 구조 - 배포용 문서
Hwp Document File Formats -
Document for Distribution

revision 1.2:20141009

차 례

저작권	4
본 문서에 대하여...	5
배포용 문서 데이터	6
1. 배포용 문서 데이터	6
2. 배포용 문서 데이터 복호화	7
2.1. Seed 찾기	7
2.2. 난수 배열 만들기	7
2.3. 해시코드 추출하기	8
2.4. 해시코드와 AES-128 알고리즘을 이용하여 레코드 복호화	8
변경 사항 이력	10

저작권

(주)한글과컴퓨터(이하 '한컴')는 문서 형식의 개방성과 표준화에 대하여 적극 찬성합니다. 한컴은 글 97의 문서 형식을 무상으로 지원한 바 있으며, 글 2002~2010 문서의 XML 형식은 HWPML에 대해서도 문서 형식을 공개한 바 있습니다. 개방형 문서 표준화 및 코드 관련 위원회에도 적극적으로 참여하여 파일 형식의 표준화와 개방성을 위해 노력해 왔습니다. 이러한 결과로 HWPML 스펙이 OWPML란 이름으로 한국산업표준(KS X 6101:2011)으로 제정되었습니다. 또한, 한컴오피스에서 기록물 장기보존 표준 포맷인 PDF/A-1의 지원과 ISO 국제 문서 형식인 ODF와 OOXML 파일 형식의 불러오기와 저장하기를 적극적으로 지원하였습니다.

본 문서를 열람하고자 하는 자라면 누구에게나 제공되는 것이며, 본 문서를 열람하는 것 외에 복사, 배포, 게재 및 본 문서에 기재되어 있는 내용을 사용하고자 하는 자는 한글과컴퓨터의 본 저작권을 충분히 인식하고 동의하여야 합니다.

본 문서를 누구나 열람, 복사, 배포, 게재 및 사용을 자유롭게 할 수 있습니다. 다만, 배포는 원 내용이 일체 수정되지 않은 원본 또는 복사본으로 제한됩니다. 원본 및 복사본은 한컴에서 제공하는 스펙의 최신 버전을 포함하고 있어야 합니다.

한컴은 한컴오피스 글 문서 파일(.hwp) 공개 문서에 따라 얻은 결과물을 기초로 또 다른 독점적, 배타적 권리를 취득하고 이를 (주)한글과컴퓨터를 상대로 행사하고자 하는 자를 상대로는 적극적으로 권리행사를 할 수도 있습니다.

그리고, 본 문서 및 본 문서에 기재된 내용을 참고하여 개발한 결과물에 대한 모든 저작권은 결과물을 개발한 개인 또는 단체에 있을 것입니다. 그러나 반드시 개발 결과물에 “본 제품은 한글과컴퓨터의 글 문서 파일(.hwp) 공개 문서를 참고하여 개발하였습니다.”라고 제품 내 유저인터페이스, 매뉴얼, 도움말 및 소스에 모두 기재하여야 하며 제품이 이러한 구성물이 없을 시에는 존재하는 구성물에만 기재합니다. 한컴은 본 문서 및 본 문서에 기재된 내용을 참고하여 개발한 결과물에 대해서 어떠한 정확성, 진실성도 보증하지 아니합니다.

본 문서에 대하여...

본 문서는 글 워드 프로세서의 파일 저장 형식 중, 글 2002 이후 제품에서 사용되는 글 문서 파일 형식의 배포용 문서에 관하여 설명한다.

본 문서는 글 문서 파일 형식의 배포용 문서에 관한 주요한 자료 구조에 대해서 설명한다.

글 문서 파일형식 5.0, 수식, 차트, 글 문서 파일 형식 3.0, HWPML에 관해서는 별도의 문서에서 설명한다.

배포용 문서

1. 배포용 문서 데이터

배포용 문서는 문서를 배포할 시 원본문서의 일반적인 편집을 제한하기 위해서 설계된 기능이다. 그 외에도 복사/붙이기, 인쇄를 제한하는 설정을 할 수 있다.

배포용문서와 일반문서는 다음과 같은 차이점을 가진다.

	배포용문서	일반문서
본문 스트림	ViewText/Section	BodyText/Section
암호화 여부	○	×
비고	복사, 인쇄제한을 위한 플래그를 가짐	

표 1 배포용 문서와 일반 문서의 차이점

배포용 문서는 본문 스트림을 암호화하고 있으며, 이 외에도 다음과 같은 스트림을 암호화 한다.

- ViewText/Section0 ... N
- Scripts/JScriptVersion
- Scripts/DefaultJScript
- DocHistory/HistoryLastDoc
- DocHistory/VersionLog0 ... N

해당 스트림들은 다음의 “배포용 문서 데이터” 레코드로 시작된다.

Tag ID : HWPTAG_DISTRIBUTE_DOC_DATA

자료형	길이(바이트)	설명
BYTE array[256]	256	배포용 문서 데이터
전체 길이	256	

표 2 배포용 문서 데이터

“배포용 문서 데이터” 레코드로 생성되는 데이터는 다음과 같다.

자료형	길이(바이트)	설명
WCHAR array[40]	80	- 배포용 문서를 생성할 때 사용된 암호의 SHA1 해시코드 - 암호화용 키(처음부터 16바이트까지)
WCHAR	2	배포용 문서 옵션 flag 0x01 = 복사방지 설정 0x02 = 인쇄방지 설정

전체 길이	82
-------	----

표 3 해시코드와 옵션 플래그 데이터

2. 배포용 문서 데이터 복호화

다음은 “배포용 문서 데이터”를 통해 스트림을 복호화하는 과정을 설명한다.



2.1. Seed 찾기

복호화과정에 필요한 Seed값은 “배포용 문서 데이터”의 처음 4바이트이다.

자료형	길이(바이트)	설명
UINT	4	Seed (배포용 문서 데이터의 첫 4바이트)
전체 길이	4	

표 4 Seed

2.2. 난수 배열 만들기

배포용 문서 데이터에서 해시코드를 추출하기 위해서는 우선 특정패턴을 지닌 난수배열을 생성해야 한다. 이때 패턴을 생성하기 위해 MS Visual C의 랜덤함수(srand(), rand())를 사용한다.

srand()의 초기 값은 앞서 찾은 Seed를 입력하며, rand()함수의 결과값에 따라 배열의 값이 채워진다.

배열의 값이 채워지는 방식은 다음과 같다.

1. srand() 초기화 (Seed 사용)
2. rand()함수의 결과 값을 이용하여 배열을 채운다. 단, rand()함수가 호출되는 순번에 따라 그 사용 방식이 달라진다.
 홀수번째 : 배열에 채워지는 값
 짝수번째 : 배열에 채워지는 횟수
3. 홀수번째 rand() & 0xFF의 값을 A라 하고, 짝수번째 (rand() & 0xF + 1)의 결과를 B라 할 때 배열에 A값을 B번 횟수만큼 삽입한다.
 예를 들어 A가 'a'이고, B가 3일 경우에 배열에 'a'를 3번 삽입한다.
4. 배열크기가 256이 될 때까지 3항을 반복한다.

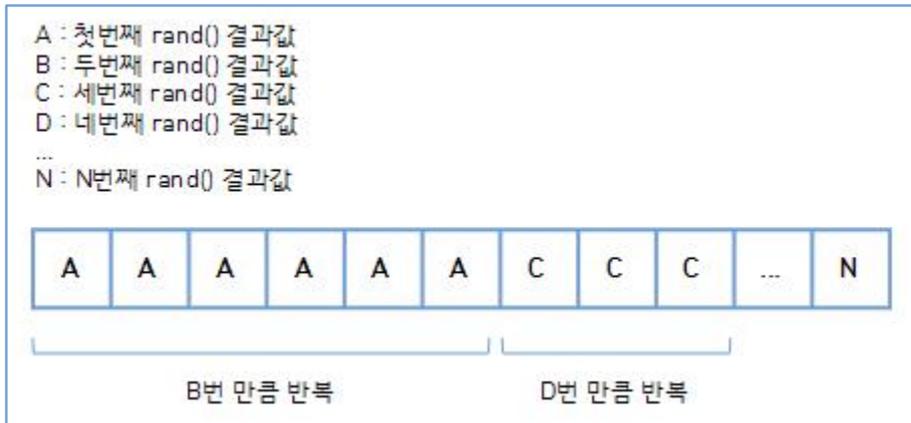


그림 4 난수 배열

2.3. 해시코드 추출하기

난수 배열을 생성한 다음에는 “배포용 문서 데이터”를 이용하여 해시코드와 옵션 플래그를 생성할 수 있다.

해시코드와 옵션 플래그는 다음의 과정을 통해 생성된다.

1. Seed으로부터 offset을 구한다. offset은 다음의 값과 같다.
 $offset = (Seed \& 0x0f) + sizeof(UINT)$
2. 난수배열과 “배포용 문서 데이터”를 XOR머지한다.
3. XOR머지된 결과값에서 offset만큼 떨어진 위치의 80바이트가 해시코드이며, 그 다음 2바이트가 옵션플래그이다.

해시코드는 배포용 문서를 만들 때 사용된 암호를 SHA1 알고리즘으로 변환한 코드이다. 해당 코드는 배포용 문서를 해제하거나 편집하는 용도로 사용된다.

2.4. 해시코드와 AES-128 알고리즘을 이용하여 레코드 복호화

해시코드의 처음 16바이트는 본문 레코드를 복호화하는 용도로 사용된다.

본문 레코드는 AES-128 ECB를 통해 암호화 되었으므로 동일 알고리즘을 이용하여 복호화할 수 있다.

복호화된 데이터는 각 레코드별 설명에 따라 읽으면 된다.

변경 사항 이력

- revision 1.2:20141009
 - 글 문서 파일 구조 파트별로 구성
 - 글 문서 파일 형식 - 배포용 문서 형식 공개

한글 문서 파일 구조 - 배포용 문서

발행처 (주) 한글과컴퓨터
주 소 (우) 463-400
경기도 성남시 분당구 대왕판교로
644번길 49 한컴타워 10층
전화: (031) 627-7000
팩스: (031) 627-7709
