

3. 한글 파일 구조

3.1. 한글 파일 구조 요약

한글의 문서 파일은 개괄적으로 다음 표와 같은 구조를 가진다. 복합 파일(Compound File) 구조를 가지기 때문에, 내부적으로 스토리지(Storage)와 스트림(Stream)을 구별하기 위한 이름을 가진다. 하나의 스트림에는 일반적인 바이너리나 레코드 구조로 데이터가 저장되고, 스트림에 따라서 압축/암호화되기도 한다.

 Storage

 Stream

설명	구별 이름	길이(바이트)	레코드 구조	압축/암호화
파일 인식 정보	 FileHeader	고정		
문서 정보	 DocInfo	고정	✓	✓
본문	 BodyText	가변	✓	✓
	 Section0			
	 Section1			
	 ...			
문서 요약	 \005HwpSummaryInformation	고정		
바이너리 데이터	 BinData	가변		✓
	 BinaryData0			
	 BinaryData1			
	 ...			
미리보기 텍스트	 PrvText	고정		
미리보기 이미지	 PrvImage	가변		
문서 옵션	 DocOptions	가변		
	 _LinkDoc			
	 DrmLicense			
	 ...			
스크립트	 Scripts	가변		
	 DefaultJScript			
	 JScriptVersion			
	 ...			
XML 템플릿	 XMLTemplate	가변		
	 Schema			
	 Instance			
	 ...			
문서 이력 관리	 DocHistory	가변	✓	✓
	 VersionLog0			
	 VersionLog1			
	 ...			

표 2 전체 구조

압축된 문서 파일의 경우 문서 파일을 읽는 쪽에서는 ‘파일 인식 정보’ 항목의 ‘압축’ 플래그를 살펴보고, 압축된 파일이면 압축을 풀어서 처리해야 한다. 이후의 설명에서는 압축이 풀린 상태의 파일을 기준으로 한다. ‘문서정보’와 ‘본문’ ‘문서 이력 관리’에 사용되는 ‘레코드 구조’는 이후 ‘데이터 레코드’란에서 구조 설명과 사용되는 레코드들에 대한 상세한 설명을 한다.

3.2. 스토리지 별 저장 정보

3.2.1. 파일 인식 정보