**DocuX 설명서**

**User API Guide (v.0.00.03.3)**

**a**

**㈜지란지교시큐리티**

목차

[1. 제품설명 1-4](#_Toc460431739)

[2. 주요기능 2-4](#_Toc460431740)

[3. 제약사항 3-5](#_Toc460431741)

[4. 지원환경 4-5](#_Toc460431742)

[5. 제품구성 5-5](#_Toc460431743)

[6. 필터 가능 포맷 6-6](#_Toc460431744)

[7. 파일포맷 코드 7-8](#_Toc460431745)

[**7.1.** **문서파일** 7-8](#_Toc460431746)

[**7.2.** **압축파일** 7-10](#_Toc460431747)

[**7.3.** **이미지파일** 7-10](#_Toc460431748)

[**7.4.** **기타 파일** 7-10](#_Toc460431749)

[8. 에러코드 8-12](#_Toc460431750)

[9. 실행 파일 바이너리 9-13](#_Toc460431751)

[**9.1.** **Unix/Linux 계열 실행방법** 9-13](#_Toc460431752)

[**9.2.** **Windows 계열 실행방법** 9-14](#_Toc460431753)

[**9.3.** **실행옵션** 9-15](#_Toc460431754)

[10. API 가이드 10-17](#_Toc460431755)

[**10.1.** **인코딩 Info** 10-17](#_Toc460431756)

[**10.2.** **Filter 옵션코드** 10-17](#_Toc460431757)

[**10.3.** **구조체 Info** 10-17](#_Toc460431758)

[**10.4.** **Function Info** 10-18](#_Toc460431759)

# **제품설명**

DocuX 는 다양한 문서 내의 텍스트 추출 기능을 제공합니다.

타 제품 내에 요소 기능을 처리하기 위한 목적으로 사용될 수 있으며, 아래 기술되어 있는 두 가지 방식 중 택일하여 기능을 구현할 수 있습니다.

요소 기능 예) 개인 정보 검출, 문서 내용 검색 외

**기술문의** : ㈜지란지교시큐리티 연구소 개발2팀 이상현

**제품문의** : ㈜지란지교시큐리티 연구소 개발2팀 조현주

# **주요기능**

* 파일 내용에 의한 파일 포맷 인식
* 문서 개요 정보 추출
* 압축파일 내 파일 필터링 지원
* 중첩된 압축파일 필터링 지원
* 메일 첨부파일 필터링 지원
* 문서 내에 포함된 파일 필터링 지원
* 다국어 인코딩 지원
* 파일/메모리 Output 인터페이스 지원

# **제약사항**

파일에 기록된 모든 데이터를 추출할 수 없으며 "필터 가능 파일 포맷" 표에 명시된 "추출 내용"에 한해 텍스트를 추출합니다.

추출된 텍스트는 원문에 기록된 어떠한 서식 정보(글꼴, 단락 등) 혹은 구조 정보(표 , 머리글, 바닥글, 주석 등)도 포함하지 않으며 추출되는 순서 또한 화면의 것과 다를 수 있습니다.

실행 파일 및 라이브러리 사용 도중 가능성은 매우 적으나 Segmentation Fault, 무한 루프 등의 Crash 현상이 발생할 수 있으므로 적절한 대응 방안을 마련하여 사용해야 합니다

# **지원환경**

* Linux 계열 (CentOS, Red Hat, Fedora, Ubuntu, FreeBSD, Suse 등)
* Unix 계열 (AIX, HP-UX, Solaris)
* Windows 95, 98, 2000, XP, 2003, 2008, Vista, 7, 10

# **제품구성**

제공되는 환경에 따라 파일 구성이 다릅니다.

|  |  |  |  |
| --- | --- | --- | --- |
| **구 분** | **파 일 명** | **설 명** | **비 고** |
| Unix/Linux 계열 | libdocux.a | DocuX 정적 라이브러리 |  |
| docux.h | DocuX 헤더파일 |  |
| docux | DocuX Unix/Linux 계열 실행 바이너리 | Stand Alone |
| Windows 계열 | libdocux.lib | DocuX DLL BUILD용 정적라이브러리 |  |
| libdocux.dll | DocuX 동적 라이브러리 |  |
| docux.h | DocuX 헤더파일 |  |
| docux.exe | DocuX Windows 계열 실행 바이너리 | Stand Alone |

# **필터 가능 포맷**

| **파일 포맷** | **버전** | **대표 확장자** |
| --- | --- | --- |
| Zip Archive Data |  | zip |
| RAR Archive Data |  | rar |
| Gzip Compressed Data |  | gz |
| Bzip Compressed Data |  | bz2 |
| Tar Archive Data |  | tar |
| 7zip Compressed Data |  | 7z |
| ESTsoft ALZIP Compressed File |  | alz |
| ESTsoft EGG Compressed File |  | egg |
| Plain Text File |  | txt |
| Rich Text Format Document |  | rtf |
| HyperText Markup Language Document |  | htm, html |
| Extensible Markup Language Document |  | xml |
| MIME HTML |  | eml, mht |
| Portable Document Format File | 1.0 ~ 1.7 | pdf |
| 핸디 아리랑 |  | hwd |
| Handysoft approval document |  | hwx |
| Handysoft document |  | hwn |
| 아래아 한글 3 | 2.x  3.x  95  97(815) | hwp |
| 아래아 한글 2000 | 워디안  2002  2002 SE  2004  2005  2007  2010  2014 | hwp, hwpx |
| 아래아 한글 XML |  | hml, hwpml |
| Microsoft Office Word | 96  97  2000  2003 | doc |
| Microsoft Office Presentation | 95  97  2000  2003 | ppt |
| Microsoft Office Spreadsheet | 96  97  2000  2003 | xls |
| Microsoft Office Word 2007 | 2007  2010  2013 | docx, doc |
| Microsoft Office Presentation 2007 | 2007  2010  2013 | pptx, ppt |
| Microsoft Office Spreadsheet 2007 | 2007  2010  2013 | xlsx, xls |
| MS Outlook Message | 2000  2003  2007  2010  2013 | msg |
| Microsoft compressed HTML help files | v1 ~ v4 | chm |
| Open Office word | 1.0 ~ 3.0 | sxw, odt, ott |
| Open Office Calc | 1.0 ~ 3.0 | sxc, ods, stc |
| Open Office Impress | 1.0 ~ 3.0 | sxi, odp, stl |
| Wordperfect | 4.x ~ 14.x | wpd |

# **파일포맷 코드**

## **문서파일**

| **포맷 이름** | **코드 값** | **포맷 설명** |
| --- | --- | --- |
| DOCUX\_FORMAT\_HWP20 | 100101 | Hangul Word-Processor Document, version 2 |
| DOCUX\_FORMAT\_HWP21 | 100102 | Hangul Word-Processor Document, version 2 |
| DOCUX\_FORMAT\_HWP30 | 100103 | Hangul Word-Processor Document, version 3 |
| DOCUX\_FORMAT\_HWP50 | 100104 | Hangul Word-Processor Document, version 5 |
| DOCUX\_FORMAT\_HWPML | 100105 | Hangul Word-Processor Markup Language |
| DOCUX\_FORMAT\_ARIRANG | 100201 | Handysoft Arirang Document |
| DOCUX\_FORMAT\_HWN | 100203 | Hwn for Handysoft Groupware |
| DOCUX\_FORMAT\_HWX | 100204 | Hwx for Handysoft Groupware |
| DOCUX\_FORMAT\_MSWORD80 | 100401 | Microsoft Word Document 97 |
| DOCUX\_FORMAT\_MSWORD70 | 100402 | Microsoft Word Document 95 |
| DOCUX\_FORMAT\_MSWORD60 | 100403 | Microsoft Word Document 95 |
| DOCUX\_FORMAT\_MSWORD95 | 100404 | Microsoft Word Document 95 |
| DOCUX\_FORMAT\_XLS97 | 100501 | Microsoft Excel Spreadsheet 97 |
| DOCUX\_FORMAT\_XLS95 | 100502 | Microsoft Excel Spreadsheet 95 |
| DOCUX\_FORMAT\_XLS40 | 100503 | Microsoft Excel Spreadsheet 40 |
| DOCUX\_FORMAT\_PPT97 | 100601 | Microsoft PowerPoint Document 97 |
| DOCUX\_FORMAT\_PPT95 | 100602 | Microsoft PowerPoint Document 95 |
| DOCUX\_FORMAT\_PDF10 | 100701 | Portable Document Format(PDF) V1.0 |
| DOCUX\_FORMAT\_PDF11 | 100702 | Portable Document Format(PDF) V1.1 |
| DOCUX\_FORMAT\_PDF12 | 100703 | Portable Document Format(PDF) V1.2 |
| DOCUX\_FORMAT\_PDF13 | 100704 | Portable Document Format(PDF) V1.3 |
| DOCUX\_FORMAT\_PDF14 | 100705 | Portable Document Format(PDF) V1.4 |
| DOCUX\_FORMAT\_PDF15 | 100706 | Portable Document Format(PDF) V1.5 |
| DOCUX\_FORMAT\_PDF16 | 100707 | Portable Document Format(PDF) V1.6 |
| DOCUX\_FORMAT\_PDF17 | 100708 | Portable Document Format(PDF) V1.7 |
| DOCUX\_FORMAT\_DWG,\_R2\_5 | 100801 | AutoCAD Release 2.50 |
| DOCUX\_FORMAT\_DWG,\_R2\_6 | 100802 | AutoCAD Release 2.60 |
| DOCUX\_FORMAT\_DWG,\_R9 | 100803 | AutoCAD Release 9 |
| DOCUX\_FORMAT\_DWG,\_R10 | 100804 | AutoCAD Release 10 |
| DOCUX\_FORMAT\_DWG,\_R11 | 100805 | AutoCAD Release 11 or 12 |
| DOCUX\_FORMAT\_DWG,\_R13\_SUB10 | 100806 | AutoCAD Release 13 (Sub type 10) |
| DOCUX\_FORMAT\_DWG,\_R13\_SUB11 | 100807 | AutoCAD Release 13 (Sub type 11) |
| DOCUX\_FORMAT\_DWG,\_R13\_SUB12 | 100808 | AutoCAD Release 13 (Sub type 12) |
| DOCUX\_FORMAT\_DWG,\_R14\_SUB13 | 100809 | AutoCAD Release 14 (Sub type 30) |
| DOCUX\_FORMAT\_DWG,\_R14\_SUB14 | 100810 | AutoCAD Release 14 (Sub type 14) |
| DOCUX\_FORMAT\_DWG,\_R2000 | 100811 | AutoCAD Release 2000 |
| DOCUX\_FORMAT\_DWG,\_R2004 | 100812 | AutoCAD Release 2004 |
| DOCUX\_FORMAT\_DWG,\_R2007 | 100813 | AutoCAD Release 2007 |
| DOCUX\_FORMAT\_DWG,\_R2010 | 100814 | AutoCAD Release 2010 |
| DOCUX\_FORMAT\_DWG,\_R2013 | 100815 | AutoCAD Release 2013 |
| DOCUX\_FORMAT\_HTML | 100900 | HyperText Markup Language Document |
| DOCUX\_FORMAT\_XML | 101000 | Extensible Markup Language Document |
| DOCUX\_FORMAT\_MSG | 101100 | MS Outlook Message |
| DOCUX\_FORMAT\_MHT | 101200 | MHTML |
| DOCUX\_FORMAT\_EML, | 101300 | Common File Extension for E-mail File |
| DOCUX\_FORMAT\_CHM, | 101400 | Microsoft compressed HTML help files |
| DOCUX\_FORMAT\_WORDPERFECT | 101500 | WordPerfect Document file |
| DOCUX\_FORMAT\_RTF | 101600 | Rich Text Format Document |
| DOCUX\_FORMAT\_WKS | 101700 | WKS |
| DOCUX\_FORMAT\_TXT | 101800 | Plain Text File |
| DOCUX\_FORMAT\_OFFICE2007\_DOCX, | 101901 | Microsoft Office 2007 Word XML |
| DOCUX\_FORMAT\_OFFICE2007\_PPTX, | 101902 | Microsoft Office 2007 Presentation XML |
| DOCUX\_FORMAT\_OFFICE2007\_XLSX, | 101903 | Microsoft Office 2007 Spreadsheet XML |
| DOCUX\_FORMAT\_OFFICE2007\_PASSWORD | 101904 | Microsoft Office 2007 XML Protect Password |
| DOCUX\_FORMAT\_OPENOFFICE | 102000 | Open Office File Format |
| DOCUX\_FORMAT\_OPENOFFICE\_WORD | 102001 | Open Office Word Document |
| DOCUX\_FORMAT\_OPENOFFICE\_SPREADSHEET | 102002 | Open Office Spreadsheet |
| DOCUX\_FORMAT\_OPENOFFICE\_PRESENTATION | 102003 | Open Office Presentation |
| DOCUX\_FORMAT\_OPENOFFICE\_DATABASE | 102004 | Open Office Database |
| DOCUX\_FORMAT\_OPENOFFICE\_DRAWING | 102005 | Open Office Drawing |
| DOCUX\_FORMAT\_OPENOFFICE\_ORG\_WORD | 102006 | OpenOffice.org XML Word Document |
| DOCUX\_FORMAT\_OPENOFFICE\_ORG\_SPREADSHEET | 102007 | OpenOffice.org XML Spreadsheet |
| DOCUX\_FORMAT\_OPENOFFICE\_ORG\_PRESENTATION | 102008 | OpenOffice.org XML Presentation |
| DOCUX\_FORMAT\_XHTML | 102100 | Extensible Hypertext Markup Language Document |

## **압축파일**

|  |  |  |
| --- | --- | --- |
| **포맷 코드** | **코드 값** | **설명** |
| DOCUX\_FORMAT\_ZIP , | 200100 | Zip Archive Data |
| DOCUX\_FORMAT\_RAR, | 200200 | RAR Archive Data |
| DOCUX\_FORMAT\_GZIP, | 200300 | Gzip Compressed Data |
| DOCUX\_FORMAT\_BZIP2, | 200400 | Bzip2 Compressed Data |
| DOCUX\_FORMAT\_Z, | 200500 | Z Compressed Data |
| DOCUX\_FORMAT\_TAR, | 200600 | Tar Archive Data |
| DOCUX\_FORMAT\_7Z, | 200800 | 7zip Compressed Data |
| DOCUX\_FORMAT\_ALZ, | 200900 | ESTsoft ALZIP Compressed File |
| DOCUX\_FORMAT\_EGG, | 201000 | ESTsoft EGG Compressed File |

## **이미지파일**

|  |  |  |
| --- | --- | --- |
| **포맷 코드** | **코드 값** | **설명** |
| DOCUX\_FORMAT\_BMP | 300100 | Bitmap Image File |
| DOCUX\_FORMAT\_PS | 300200 | PostScript |
| DOCUX\_FORMAT\_PCX | 300300 | ZSOFT Paintbrush File |
| DOCUX\_FORMAT\_JPEG | 300400 | Jpeg Image File |
| DOCUX\_FORMAT\_TIFF | 300500 | Tagged Image File Format |
| DOCUX\_FORMAT\_GIF | 300600 | Graphics Interchange Format File |
| DOCUX\_FORMAT\_PIC | 300700 | Pictor Picture Image Format File |
| DOCUX\_FORMAT\_PNG | 300800 | Portable Network Graphics file |

*\* 이미지 포맷은 포맷정보 검출만 지원합니다.*

## **기타 파일**

|  |  |  |
| --- | --- | --- |
| **포맷 코드** | **코드 값** | **설명** |
| DOCUX\_FORMAT\_JAVA\_CLASS | 400100 | Compiled Java Class File |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM11 | 400101 | Compiled Java Class File By JVM 1.1 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM12 | 400102 | Compiled Java Class File By JVM 1.2 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM13 | 400103 | Compiled Java Class File By JVM 1.3 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM14 | 400104 | Compiled Java Class File By JVM 1.4 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM15 | 400105 | Compiled Java Class File By JVM 1.5 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM16 | 400106 | Compiled Java Class File By JVM 1.6 |
| DOCUX\_FORMAT\_JAVA\_CLASS\_JVM17 | 400107 | Compiled Java Class File By JVM 1.7 |

*\* Etc 포맷은 포맷정보 검출만 지원합니다.*

# **에러코드**

| **포맷 코드** | **코드 값** | **설명** |
| --- | --- | --- |
| DOCUX\_INIT\_CONF\_ERR | 10 | 초기 설정이 되지않았습니다 |
| DOCUX\_INTERNAL\_ERR, | 11 | 내부 처리 문제가 발생하였습니다 |
| DOCUX\_MEMORY\_ERR, | 12 | 메모리 에러가 발생되었습니다 |
| DOCUX\_DISK\_ERR, | 13 | 디스크 에러가 발생되었습니다 |
| DOCUX\_HAVE\_PASSWORD, | 14 | 대상파일에 암호가 존재합니다 |
| DOCUX\_BROKEN\_FILE, | 15 | 손상된 파일입니다 |
| DOCUX\_INVALID\_PARAM, | 16 | 유효하지 않은 파라미터입니다 |
| DOCUX\_FILE\_NOT\_EXIST | 30 | 파일이 존재하지 않습니다 |
| DOCUX\_FILE\_OPEN\_ERR | 31 | 파일 열기 오류가 발생되었습니다 |
| DOCUX\_FILE\_READ\_ERR | 32 | 파일 읽기 오류가 발생되었습니다 |
| DOCUX\_FILE\_WRITE\_ERR | 33 | 파일 쓰기 오류가 발생되었습니다 |
| DOCUX\_FILE\_NAME\_NULL | 34 | 파일 이름이 입력되지 않았습니다 |
| DOCUX\_FILE\_NOT\_REGULAR | 35 | 일반파일이 아닙니다(Directory 등) |
| DOCUX\_FILE\_SIZE\_ERR | 36 | 파일size가 유효하지 않습니다(0Byte 등) |
| DOCUX\_LICENSE\_KEY\_ERR | 50 | 라이선스 키에 오류가 있습니다 |
| DOCUX\_LICENSE\_EXPIRATION\_ERR | 51 | 라이선스가 만료되었습니다 |
| DOCUX\_DEMO\_EXPIRATION | 52 | 데모 라이선스가 만료되었습니다 |
| DOCUX\_UNSUPPORT\_VERSION | 53 | 지원하지 않는 버전입니다 |
| DOCUX\_UNSUPPORT\_FORMAT | 255 | 지원하지 않는 포맷입니다 |

# **실행 파일 바이너리**

## **Unix/Linux 계열 실행방법**

**$ ./docux [Options] <Source File> [Destination File]**

1. *X-Window 의 Terminal을 실행하거나 네트웍 접속으로 shell을 엽니다.*
2. *docux 실행 파일이 있는 디렉토리로 이동합니다.*

* *$ cd /home/user/docux/bin/*

1. *다음 명령으로 해당 파일의 텍스트를 추출합니다. (Default 출력인코딩 : UTF-8)*

* *$ ./docux /sample\_data/sample.doc (터미널 출력 – Default 출력인코딩)*
* *$ ./docux –o euc-kr /sample\_data/sample.doc (터미널 출력 – EUC-KR)*
* *$ ./docux /sample\_data/sample.doc sample.txt (파일 출력 - Default 출력인코딩)*
* *$ ./docux –o euc-kr /sample\_data/sample.doc sample.txt (파일 출력 – EUC-KR)*

1. *화면 또는 파일에 출력되는 결과를 확인합니다.*

*docux 수행시 libstdc++.so, libgcc.so 가 설치되지 않았거나 경로를 찾지 못할 경우 구동이 되지 않을 수 있습니다.*

*문서 필터가 빌드된 gcc 혹은 기타 C/C++ 컴파일러의 버전에 맞는 stdc++, c++ shared object를 시스템에 설치하거나 LD\_LIBRARY\_PATH과 같은 환경 변수 설정으로 해당 라이브러리 파일이 인식되도록 환경을 구성하세요*.

## **Windows 계열 실행방법**

**C:\>docux.exe [Options] <Source File> [Destination File]**

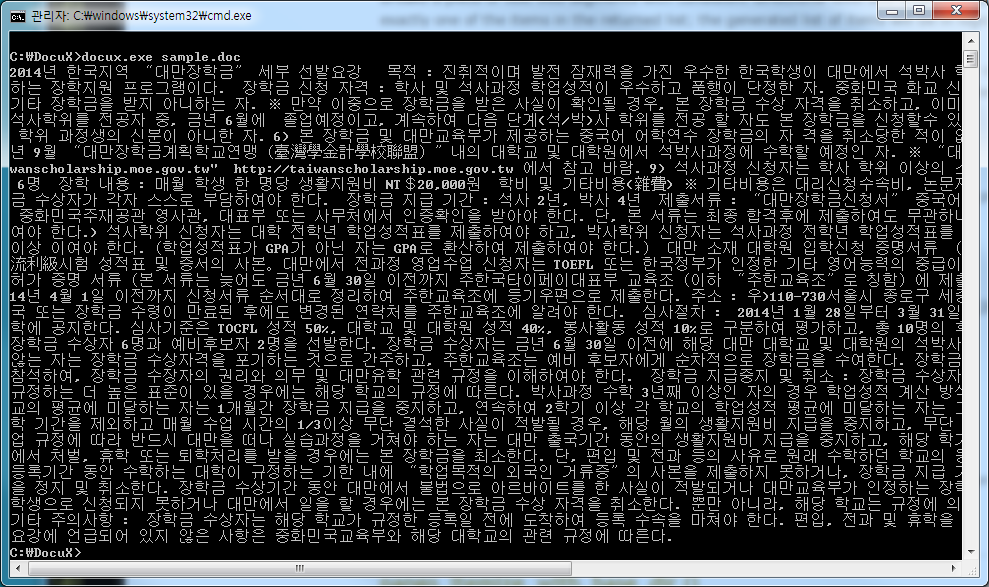
1. *Command Prompt 창을 엽니다*
2. *docux 실행 파일이 있는 디렉토리로 이동합니다.*

* *C:\> cd DocuX*
* *C;\Docux>*

1. *다음 명령으로 해당 파일의 텍스트를 추출합니다. (Default 출력인코딩 : UTF-8)*

* *C:\DocuX> docux.exe sample.doc (터미널 출력 – Default 출력인코딩)*
* *C:\DocuX> docux.exe –o euc-kr sample.doc (터미널 출력 – EUC-KR)*
* *C:\DocuX> docux.exe sample.doc sample.txt (파일 출력 - Default 출력인코딩)*
* *C:\DocuX> docux.exe –o euc-kr sample.doc sample.txt (파일 출력 – EUC-KR)*

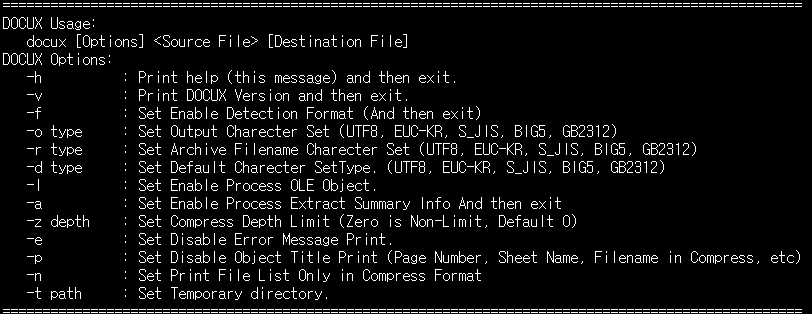
1. *화면 또는 파일에 출력되는 결과를 확인합니다.*



**SAMPLE**

## **실행옵션**

|  |  |  |
| --- | --- | --- |
| **옵션 코드** | **기능** | **비고** |
| -h | Help Message 출력 | 해당기능 수행 후 프로그램 종료 |
| -v | Version Info 출력 |
| -f | 문서 포맷정보 출력 |
| -i Encoding Type | 입력 인코딩 강제 지정 (대소문자 무관) | UTF8, EUC-KR, S\_JIS, BIG5, GB2312 |
| -o Encoding Type | 출력 인코딩 지정 (대소문자 무관) | UTF8, EUC-KR, S\_JIS, BIG5, GB2312  (옵션 미지정 시 항목별 기본 인코딩 Type 적용됨,  -o : UTF-8, -r : EUC-KR,-d : EUC-KR) |
| -r Encoding Type | Archive 내 파일이름 인코딩 지정  **(Unicode 파일이름은 적용 안됨)** |
| -d Encoding Type | 인코딩 검출 실패 시 사용될 인코딩 지정 |
| -l | OLE 추출 Enable |  |
| -a | 문서 속성정보 출력 | 해당기능 수행 후 종료 |
| -z depth\_num | 중첩된 압축파일의 처리 depth를 제한 | 0(Zero) : Non-limit (default)  positive number : Limit Depth |
| -e | 에러정보(결과코드&설명) 출력 제거 | Default : 에러정보 출력 |
| -p | Object Title 출력 제거 | PDF : Page Number  PPT&PPTX : Page Number  XLS&XLSX : Sheet Name  Compress File : Compressed Filename |
| -n | 압축파일에서 파일이름만 출력 | 내용은 추출하지 않음 |
| -t path | DocuX 내부 사용 임시파일 경로 지정 | 미지정(또는 잘못된 입력)시 현재 작업Path 이용  드라이브명(Windows계열) 또는 절대경로(Unix계열) 미지정시 Default 드라이브명(C:\) 또는 절대경로 강제적용 |

**

***주의)*** *옵션코드 형식에 맞지 않게 입력 될 경우 일부 옵션이 적용되지 않거나 정상적으로 동작하지 않을 수 있습니다.*

# **API 가이드**

## **인코딩 Info**

|  |  |  |
| --- | --- | --- |
| **인코딩 코드** | **Value** | **설명** |
| **JSDX\_EUC\_KR** | 949 | 한글 완성형 |
| **JSDX\_SJIS** | 932 | Shift JIS |
| **JSDX\_EUC\_JP** | 20932 | Three Japanese character set standards |
| **JSDX\_BIG5** | 950 | 중문 번체 |
| **JSDX\_GB2312** | 936 | 중문 간체 |
| **JSDX\_UTF8** | 1 | UCS Transformation Format 8-bit |

## **Filter 옵션코드**

|  |  |  |
| --- | --- | --- |
| **Filter 옵션** | **Code** | **설명** |
| **JSDX\_OPT\_NONE** | 0x00000000 | 필터옵션 없음(단독사용) |
| **JSDX\_OPT\_ARCHIVE\_FILELIST\_ONLY** | 0x00000001 | 압축파일 내 파일이름만 출력(내용X) |
| **JSDX\_OPT\_ARCHIVE\_SHOW\_PATH** | 0x00000002 | 압축 내 파일명 출력 시 Full Path 출력 |
|  |  |  |
|  |  |  |
|  |  |  |

## **구조체 Info**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **구조체 이름** | **자료형** | **맴버** | **설명** | **비고** |
| **JSDX\_CONF\_OPT\_ST** | int | limit\_buffer\_size | 출력버퍼 size limit  (byte 단위) | 미지원 |
| int | force\_input\_encoding | 입력 인코딩 지정 | 인코딩 Info 및  실행옵션 참조 |
| int | force\_output\_encoding | 출력 인코딩 지정 |
| int | force\_archive\_encoding | 압축 내 파일이름 인코딩 지정 |
| int | force\_default\_encoding | 인코딩 자동검출 실패 시 사용될 기본인코딩 지정 |
| int | proc\_ole\_object | Ole 추출여부 지정 | (On:1/Off:0) |
| int | limit\_depth | 중첨된 압축파일  처리범위(Depth) 지정 | 음수 : 기본값  (Unlimit)  양수 :  limit depth |
| int | object\_title\_print | PageNumber, SheetName,  Compressed Filename  출력 유무지정 | (On:0/Off:1) |
| Unsigned  long int | filter\_option | 필터에 적용될 옵션 지정 | 옵션코드참조  (중복지정가능) |
| char \* | Inner\_temp\_dir | 내부 임시파일 경로 지정  (절대경로 사용) | 미지정시 현재 작업경로 사용 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **구조체 이름** | **자료형** | **맴버** | **설명** | **비고** |
| **JSDX\_SummaryInfo** | int | format\_code | 포맷타입 코드 | 지원포맷참조 |
| int | codepage | MBCS Code Page Value  (Default: 0 – Non Set) | 인코딩 Info  참조 |
| char \* | info\_title | 타이틀 |  |
| char \* | info\_subject | 주제 |  |
| char \* | info\_author | 저자 |  |
| char \* | info\_keywords | 키워드 |  |
| char \* | info\_comments | 설명 |  |
| char \* | info\_template | 문서 템플릿 |  |
| char \* | info\_last\_author | 최종 수정자 |  |
| char \* | info\_rev\_no | 문서 버전 |  |
| char \* | info\_app\_name | 프로그램명 |  |
| char \* | info\_create\_time | 문서 생성일자 |  |
| char \* | info\_save\_time | 문서 최근 수정일자 |  |

## **Function Info**

|  |  |
| --- | --- |
| **JSDX\_InitConfig** | |
| **기능** | 필터링 환경을 설정함 |
| **원형** | int JSDX\_InitConfig(char \*license\_key, unsigned long int reserved); |
| **Parameters** | license\_key  - 라이선스 키  reserved  - reserved |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * 기본 출력 인코딩은 UTF-8로 지정되어 있으며, 다른 인코딩 타입을 사용하기   위해서는 JSDX\_InitConfigEx() 함수 사용 |
| **Example** | int ret;  if((ret = JSDX\_InitConfig("DOCUX", 0) != DOCUX\_OK)  {  printf("filter error %d\n", ret);  return ret;  } |
| **See Also** | JSDX\_InitConfigEx |

|  |  |
| --- | --- |
| **JSDX\_InitConfigEx** | |
| **기능** | 추가 옵션을 포함한 필터링 환경을 설정함 |
| **원형** | int JSDX\_InitConfigEx(char \*license\_key, unsigned long int reserved,  JSDX\_CONF\_OPT\_ST opt); |
| **Parameters** | license\_key  - 라이선스 키  reserved  - reserved  opt  - 추가 옵션 정보 (구조체 Info 참조) |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * 기본 출력 인코딩은 UTF-8로 지정되어 있으며, 다른 인코딩 타입으로   변환하기 위해서는 docux\_opt.force\_output\_encoding 옵션을 설정해야함  (인코딩 Info 참조) |
| **Example** | int ret;  JSDX\_CONF\_OPT\_ST docux\_opt;  memset(&docux\_opt, 0x00, sizeof(JSDX\_CONF\_OPT\_ST));  docux\_opt.force\_input\_encoding = 0;  docux\_opt.force\_output\_encoding = EUC-KR; (Zero is Default Encoding - UTF-8)  docux\_opt.proc\_ole\_object = 0; // (On : 1, Off : 0)  docux\_opt.limit\_depth = 0; // (0 : Non-Limit, etc : Limit depth)  docux\_opt.object\_title\_print = 0; // (On : 0, Off : 1)  //docux\_opt.filter\_option = (JSDX\_OPT\_NONE); // 옵션미적용  docux\_opt.filter\_option = (JSDX\_OPT\_ARCHIVE\_FILELIST\_ONLY | JSDX\_OPT\_ARCHIVE\_SHOW\_PATH); // 파일리스트만 출력 및 Full Path 출력 옵션적용  if((ret = JSDX\_InitConfigEx("DOCUX", 0, docux\_opt)) != DOCUX\_OK)  {  printf("filter error %d\n", ret);  return ret;  } |
| **See Also** | JSDX\_InitConfig |

|  |  |
| --- | --- |
| **JSDX\_ExecuteDocux** | |
| **기능** | 파일로부터 텍스트를 추출하고 결과파일(or 터미널)에 기록함 |
| **원형** | int JSDX\_ExecuteDocux(char \*src, char \*dest); |
| **Parameters** | src  - 텍스트를 추출할 대상 파일 경로  Dest  - 결과를 저장할 파일 경로(미입력 시 표준출력으로 추출) |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * dest 파일에 추출결과 또는 에러코드가 기록됩니다. * 출력되는 결과는 지정된 출력 인코딩에 맞춰서 저장됩니다. * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx) |
| **Example** | int ret;  ret = JSDX\_ExecuteDocux("/home/src/sample.data", “/home/dest/result.txt”);  if(ret != DOCUX\_OK )  {  printf("filter error %d\n", ret);  return ret;  }  printf("succeed\n");  return DOCUX\_OK; |
| **See Also** |  |

|  |  |
| --- | --- |
| **JSDX\_ExecuteDocux4Unicode** | |
| **기능** | 파일로부터 텍스트를 추출하고 결과파일(or 터미널)에 기록함 |
| **원형** | int JSDX\_ExecuteDocux(wchar\_t \*src, char \*dest); |
| **Parameters** | src  - 텍스트를 추출할 대상 파일 경로(UTF16 인코딩 파일명)  Dest  - 결과를 저장할 파일 경로(미입력 시 표준출력으로 추출) |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * dest 파일에 추출결과 또는 에러코드가 기록됩니다. * 출력되는 결과는 지정된 출력 인코딩에 맞춰서 저장됩니다. * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx)   * Dest 파일명은 유니코드를 지원하지 않습니다. * Windows 버전 전용 API |
| **Example** | int ret;  ret = JSDX\_ExecuteDocux(L"/home/src/sample.data", “/home/dest/result.txt”);  if(ret != DOCUX\_OK )  {  printf("filter error %d\n", ret);  return ret;  }  printf("succeed\n");  return DOCUX\_OK; |
| **See Also** | JSDX\_ExecuteDocux |

|  |  |
| --- | --- |
| **JSDX\_ExecuteDocuxMem** | |
| **기능** | 파일로부터 텍스트를 추출하고 메모리에 기록함 |
| **원형** | int JSDX\_ExecuteDocuxMem(char \*src, char \*\*dest\_buffer, int \*dest\_len); |
| **Parameters** | src  - 텍스트를 추출할 대상 파일 경로  dest\_buffer  - 추출결과가 저장된 메모리의 주소를 저장할 문자열 포인터 주소  dest\_len  - 추출결과의 사이즈를 기록할 int형 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * dest\_buffer에 추출결과 또는 에러코드가 기록됩니다. * 사용 후 dest\_buffer 메모리를 해제하여야 합니다.   (dest\_buffer 의 할당 크기는 dest\_len을 참조)   * 정상동작 후 결과가 없을 경우 메모리에 Null String을 반환합니다(‘\0’) * 출력되는 결과는 지정된 출력 인코딩에 맞춰서 메모리에 저장되며,   Default 출력 인코딩은 UTF-8입니다. (MBCS 출력의 경우 결과 확인을 위해서는  별도의 처리(MBCS 인코딩 출력기능 등)가 필요합니다.)   * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx) |
| **Example** | int ret;  int len = 0;  char \*dest\_buffer = NULL;  ret = JSDX\_ExecuteDocuxMem("/home/src/sample.data",&dest\_buffer, &len);  if(ret != DOCUX\_OK )  {  printf("filter error %d\n", ret);  JSDX\_MemRelease(dest\_buffer);  return ret;  }  printf("succeed\n");  JSDX\_MemRelease(dest\_buffer);  return DOCUX\_OK; |
| **See Also** | JSDX\_MemRelease |

|  |  |
| --- | --- |
| **JSDX\_ExecuteDocuxMem4Unicode** | |
| **기능** | 파일로부터 텍스트를 추출하고 메모리에 기록함 |
| **원형** | int JSDX\_ExecuteDocuxMem(wchar\_t \*src, char \*\*dest\_buffer, int \*dest\_len); |
| **Parameters** | src  - 텍스트를 추출할 대상 파일 경로 (UTF16 인코딩 파일명)  dest\_buffer  - 추출결과가 저장된 메모리의 주소를 저장할 문자열 포인터 주소  dest\_len  - 추출결과의 사이즈를 기록할 int형 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * dest\_buffer에 추출결과 또는 에러코드가 기록됩니다. * 사용 후 dest\_buffer 메모리를 해제하여야 합니다.   (dest\_buffer 의 할당 크기는 dest\_len을 참조)   * 정상동작 후 결과가 없을 경우 메모리에 Null String을 반환합니다(‘\0’) * 출력되는 결과는 지정된 출력 인코딩에 맞춰서 메모리에 저장되며,   Default 출력 인코딩은 UTF-8입니다. (MBCS 출력의 경우 결과 확인을 위해서는  별도의 처리(MBCS 인코딩 출력기능 등)가 필요합니다.)   * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx)   * Windows 버전 전용 API |
| **Example** | int ret;  int len = 0;  char \*dest\_buffer = NULL;  ret = JSDX\_ExecuteDocuxMem(L"/home/src/sample.data",&dest\_buffer, &len);  if(ret != DOCUX\_OK )  {  printf("filter error %d\n", ret);  JSDX\_MemRelease(dest\_buffer);  return ret;  }  printf("succeed\n");  JSDX\_MemRelease(dest\_buffer);  return DOCUX\_OK; |
| **See Also** | JSDX\_ExecuteDocuxMem  JSDX\_MemRelease |

|  |  |
| --- | --- |
| **JSDX\_GetFormatInfo** | |
| **기능** | 대상파일의 포맷정보를 얻어옴 |
| **원형** | int JSDX\_GetFormatInfo(char \*src, int \*format\_code, char \*\*format\_name); |
| **Parameters** | src  - 대상 파일  format\_code  - 포맷코드가 저장될 int형 포인터  format\_name  - 포맷코드 설명이 저장될 char 형 포인터 주소 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * format\_name 의 크기는 문자열 길이를 통해 확인가능합니다 * 사용 후 반드시 해제되어야 합니다 * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx) |
| **Example** | int format\_code = 0;  char \*format\_name = NULL;  if((res = JSDX\_GetFormatInfo(“test.doc”, &format\_code, &format\_name)) != DOCUX\_OK)  {  // Error 처리  }  else  {  // Format Info 사용  fprintf(stdout, "[CODE] : %d\n",format\_code);  fprintf(stdout, "[NAME] : %s\n",format\_name);  }  // 사용 메모리 해제  JSDX\_MemRelease(format\_name); |
| **See Also** | JSDX\_MemRelease |

|  |  |
| --- | --- |
| **JSDX\_GetSummaryInfo** | |
| **기능** | 대상파일의 문서정보를 얻어옴 |
| **원형** | int JSDX\_GetSummaryInfo(char \*srcfn, JSDX\_SummaryInfo \*suminfo); |
| **Parameters** | srcfn  - 대상 파일  suminfo  - 문서정보가 저장될 JSDX\_SummaryInfo 구조체 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * JSDX\_SummaryInfo 내 각 항목의 메모리는 사용 후 반드시 해제되어야 합니다 * 설정 초기화를 진행해야 정상 동작합니다.   (JSDX\_InitConfig 또는 JSDX\_InitConfigEx) |
| **Example** | JSDX\_SummaryInfo summary\_info;  memset(&summary\_info, 0x00, sizeof(JSDX\_SummaryInfo)); // 초기화 필수  if((res = JSDX\_GetSummaryInfo(“test.doc”, &summary\_info)) != DOCUX\_OK)  {  // Error 처리  }  else  {  // Summary Info 사용  if(summary\_info ->info\_title != NULL)  {  Printf(“%s”, summary\_info ->info\_title);  }  …  }  // 사용후 메모리 해제  JSDX\_FreeSummaryInfo(&summary\_info); |
| **See Also** | JSDX\_FreeSummaryInfo |

|  |  |
| --- | --- |
| **JSDX\_FreeSummaryInfo** | |
| **기능** | JSDX\_SummaryInfo 구조체내 메모리를 해제 |
| **원형** | int JSDX\_FreeSummaryInfo(JSDX\_SummaryInfo \*suminfo); |
| **Parameters** | suminfo  - 문서정보가 저장된 JSDX\_SummaryInfo 구조체 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** |  |
| **Example** |  |
| **See Also** | JSDX\_GetSummaryInfo |

|  |  |
| --- | --- |
| **JSDX\_GetVersion** | |
| **기능** | 버전정보를 가져옴 |
| **원형** | int JSDX\_GetVersion(char \*version); |
| **Parameters** | version  - 버전정보를 가져올 문자열 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * x.yy.zz 형식의 버전정보 (문자열) |
| **Example** |  |
| **See Also** |  |

|  |  |
| --- | --- |
| **JSDX\_GetProgram** | |
| **기능** | 프로그램 정보를 가져옴 |
| **원형** | int JSDX\_GetProgram(char \*program); |
| **Parameters** | program  - 프로그램 정보를 가져올 문자열 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * 프로그램 이름을 가져옵니다. |
| **Example** |  |
| **See Also** |  |

|  |  |
| --- | --- |
| **JSDX\_GetLicenseDate** | |
| **기능** | License 만료 일자 정보를 가져옴 |
| **원형** | int JSDX\_GetLicenseDate(char \*date); |
| **Parameters** | date  - License 만료일 정보를 가져올 문자열 포인터 |
| **Return Value** | 성공 시  - DOCUX\_OK  실패 시  - Error Code |
| **Remarks** | * License 만료 일자 문자열 가져옵니다.   (만료일 없을 시 Null String 반환) |
| **Example** |  |
| **See Also** |  |

|  |  |
| --- | --- |
| **JSDX\_MemRelease** | |
| **기능** | 라이브러리 내에서 할당한 메모리 해제 |
| **원형** | void JSDX\_MemRelease (void \*ptr); |
| **Parameters** | ptr  - 라이브러리 내부에서 할당되어 리턴 받은 메모리 포인터 |
| **Return Value** | 없음 |
| **Remarks** |  |
| **Example** |  |
| **See Also** | JSDX\_ExecuteDocuxMem  JSDX\_GetFormatInfo |