

uCosminexus DocumentBroker Text Search Index Loader Version 3

解説書

3020-3-U72

マニュアルの購入方法

このマニュアル，および関連するマニュアルをご購入の際は，
巻末の「ソフトウェアマニュアルのサービス ご案内」をご参
照ください。

対象製品

R-1595E-33 uCosminexus DocumentBroker Text Search Index Loader Version 3 03-60 (適用 OS : Windows Server 2003 , Windows Server 2003 R2 , Windows Server 2003 R2 (x64) , Windows Server 2008 x86)

輸出時の注意

本製品を輸出される場合には、外国為替および外国貿易法ならびに米国の輸出管理関連法規などの規制をご確認の上、必要な手続きをお取りください。

なお、ご不明な場合は、弊社担当営業にお問い合わせください。

商標類

Acrobat は、Adobe Systems Incorporated (アドビシステムズ社) の商標です。

Adobe は、Adobe Systems Incorporated (アドビシステムズ社) の商標です。

CORBA は、Object Management Group が提唱する分散処理環境アーキテクチャの名称です。

GIF は、米国 CompuServe Inc. が開発したフォーマットの名称です。

Lotus は、IBM Corporation の登録商標です。

Lotus 1-2-3 は、IBM Corporation の登録商標です。

Microsoft は、米国およびその他の国における米国 Microsoft Corp. の登録商標です。

Microsoft Excel は、米国 Microsoft Corp. の商品名称です。

Microsoft Office Excel は、米国 Microsoft Corp. の商品名称です。

Microsoft Office Word は、米国 Microsoft Corp. の商品名称です。

Microsoft Word は、米国 Microsoft Corp. の商品名称です。

Microsoft, PowerPoint は、米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です。

PowerPoint は、米国およびその他の国における米国 Microsoft Corp. の登録商標です。

Visual C++ は、米国およびその他の国における米国 Microsoft Corp. の登録商標です。

Windows は、米国およびその他の国における米国 Microsoft Corp. の登録商標です。

Windows Server は、米国 Microsoft Corporation の米国およびその他の国における登録商標です。

一太郎は、(株) ジャストシステムの登録商標です。

活文, PDFstaff は、日立ソフトウェアエンジニアリング株式会社の登録商標です。

ラビニティ, libinity は、株式会社日立システムアンドサービスの登録商標です。

発行

2009 年 4 月 (第 1 版) 3020-3-U72

著作権

All Rights Reserved. Copyright (C) 2007, Hitachi, Ltd.

All Rights Reserved. Copyright (C) 2007, 2009, Hitachi Systems & Services, Ltd.

はじめに

このマニュアルは、プログラムプロダクト R-1595E-33 uCosminexus DocumentBroker Text Search Index Loader Version 3 の機能と使い方について説明したものです。

以降、このマニュアルでは、uCosminexus DocumentBroker Text Search Index Loader Version 3 のことを「テキストインデクサー」と表記します。

対象読者

このマニュアルは、テキストインデクサーを使用して、uCosminexus DocumentBroker のシステムを構築、運用および管理する方を対象としています。なお、次の内容を理解されていることを前提としています。

- uCosminexus DocumentBroker のオブジェクトモデルに関する知識
- Windows (Windows Server 2003 , Windows Server 2003 R2 , Windows Server 2008) に関する知識
- 分散オブジェクト技術に関する知識

マニュアルの構成

このマニュアルは、次に示す章と付録から構成されています。

第 1 章 テキストインデクサーの概要

テキストインデクサーの機能と処理概要について説明しています。

第 2 章 実行環境の設定

テキストインデクサーを使用するための環境設定の方法について説明しています。

第 3 章 テキストインデクサーで使用するファイル

テキストインデクサーを実行するために必要なファイルについて説明しています。

第 4 章 コマンドリファレンス

テキストインデクサーのコマンドの使用方法について説明しています。

第 5 章 プラグイン関数

テキストインデクサーが定義するプラグイン関数について説明しています。

第 6 章 トラブルシュート機能

テキストインデクサーのトラブルシュート機能とメッセージについて説明しています。

付録 A 処理対象の全文検索インデクス用プロパティ

テキストインデクサーが処理対象とする全文検索インデクス用のプロパティ種別について説明しています。

付録 B 文書空間で使用する文字コード種別が UTF-8 の場合のサポート範囲

文書空間で使用する文字コード種別が UTF-8 の場合にサポートする範囲について説明しています。

はじめに

付録 C 用語解説

このマニュアルで使用している用語について説明しています。

関連マニュアル

このマニュアルの関連マニュアルを次に示します。必要に応じてお読みください。なお、本文に記載のマニュアル名称は、「uCosminexus DocumentBroker」を「DocumentBroker」と表記しています。

uCosminexus DocumentBroker のマニュアル

uCosminexus DocumentBroker Version 3 システム導入・運用ガイド (3020-3-U71)

uCosminexus DocumentBroker を使用する環境を定義、管理および運用する場合に参照してください。

uCosminexus DocumentBroker Version 3 クラスライブラリ C++ 解説 (3000-3-F13)

uCosminexus DocumentBroker Development Kit が提供するクラスライブラリの機能と、クラスライブラリを使用するために必要なオブジェクトモデルについて知りたい場合に参照してください。

uCosminexus DocumentBroker Version 3 クラスライブラリ C++ リファレンス 基本機能編 (3000-3-F14)

クラスライブラリのクラスの詳細とメソッドの文法について知りたい場合に参照してください。

uCosminexus DocumentBroker Version 3 オブジェクト操作ツール (3000-3-F15)

クラスライブラリの API をコマンド形式で実行できるオブジェクト操作ツールについて知りたい場合に参照してください。

uCosminexus DocumentBroker Version 3 メッセージ (3000-3-F12)

uCosminexus DocumentBroker が出力するメッセージについて知りたい場合に参照してください。

関連製品のマニュアル (HiRDB)

- スケーラブルデータベースサーバ HiRDB Version 6 システム運用ガイド (UNIX(R) 用)
(3000-6-234) ¹
- スケーラブルデータベースサーバ HiRDB Version 6 システム運用ガイド (Windows(R) 用)
(3020-6-124) ¹
- スケーラブルデータベースサーバ HiRDB Version 7 システム運用ガイド (UNIX(R) 用)
(3000-6-274) ¹
- スケーラブルデータベースサーバ HiRDB Version 7 システム運用ガイド (Windows(R) 用)
(3020-6-274) ¹
- スケーラブルデータベースサーバ HiRDB Version 8 システム運用ガイド (UNIX(R) 用)
(3000-6-354) ¹
- スケーラブルデータベースサーバ HiRDB Version 8 システム運用ガイド (Windows(R) 用)

- (3020-6-354) ¹
- スケーラブルデータベースサーバ HiRDB Version 6 コマンドリファレンス (UNIX(R) 用)
- (3000-6-235) ²
- スケーラブルデータベースサーバ HiRDB Version 6 コマンドリファレンス (Windows(R) 用) (3020-6-125) ²
- スケーラブルデータベースサーバ HiRDB Version 7 コマンドリファレンス (UNIX(R) 用) (3000-6-275) ²
- スケーラブルデータベースサーバ HiRDB Version 7 コマンドリファレンス (Windows(R) 用) (3020-6-275) ²
- スケーラブルデータベースサーバ HiRDB Version 8 コマンドリファレンス (UNIX(R) 用) (3000-6-355) ²
- スケーラブルデータベースサーバ HiRDB Version 8 コマンドリファレンス (Windows(R) 用) (3020-6-355) ²
- スケーラブルデータベースサーバ HiRDB Version 6 UAP 開発ガイド (UNIX(R)/Windows(R) 用) (3000-6-236) ³
- スケーラブルデータベースサーバ HiRDB Version 6 UAP 開発ガイド (Windows(R) 用) (3020-6-126) ³
- スケーラブルデータベースサーバ HiRDB Version 7 UAP 開発ガイド (UNIX(R)/Windows(R) 用) (3000-6-276) ³
- スケーラブルデータベースサーバ HiRDB Version 7 UAP 開発ガイド (Windows(R) 用) (3020-6-276) ³
- スケーラブルデータベースサーバ HiRDB Version 8 UAP 開発ガイド (3020-6-356) ³
- スケーラブルデータベースサーバ HiRDB Version 6 メッセージ (UNIX(R)/Windows(R) 用) (3000-6-238) ⁴
- スケーラブルデータベースサーバ HiRDB Version 6 メッセージ (Windows(R) 用) (3020-6-128) ⁴
- スケーラブルデータベースサーバ HiRDB Version 7 メッセージ (UNIX(R)/Windows(R) 用) (3000-6-278) ⁴
- スケーラブルデータベースサーバ HiRDB Version 7 メッセージ (Windows(R) 用) (3020-6-278) ⁴
- スケーラブルデータベースサーバ HiRDB Version 8 メッセージ (3020-6-358) ⁴
- HiRDB 全文検索プラグイン HiRDB Text Search Plug-in Version 2 (3000-6-245) (3020-6-140) ⁵
- HiRDB 全文検索プラグイン HiRDB Text Search Plug-in Version 7 (3000-6-288) ⁵
- HiRDB 全文検索プラグイン HiRDB Text Search Plug-in Version 8 (3020-6-375) ⁵
- HiRDB Adapter for XML ユーザーズガイド (3000-6-250)

はじめに

注 1

このマニュアルでは、これらのマニュアルを「HiRDB システム運用ガイド」と表記しています。

注 2

このマニュアルでは、これらのマニュアルを「HiRDB コマンドリファレンス」と表記しています。

注 3

このマニュアルでは、これらのマニュアルを「HiRDB UAP 開発ガイド」と表記していません。

注 4

このマニュアルでは、これらのマニュアルを「HiRDB メッセージ」と表記しています。

注 5

このマニュアルでは、これらのマニュアルを「HiRDB Text Search Plug-in」と表記しています。

関連製品のマニュアル（その他）

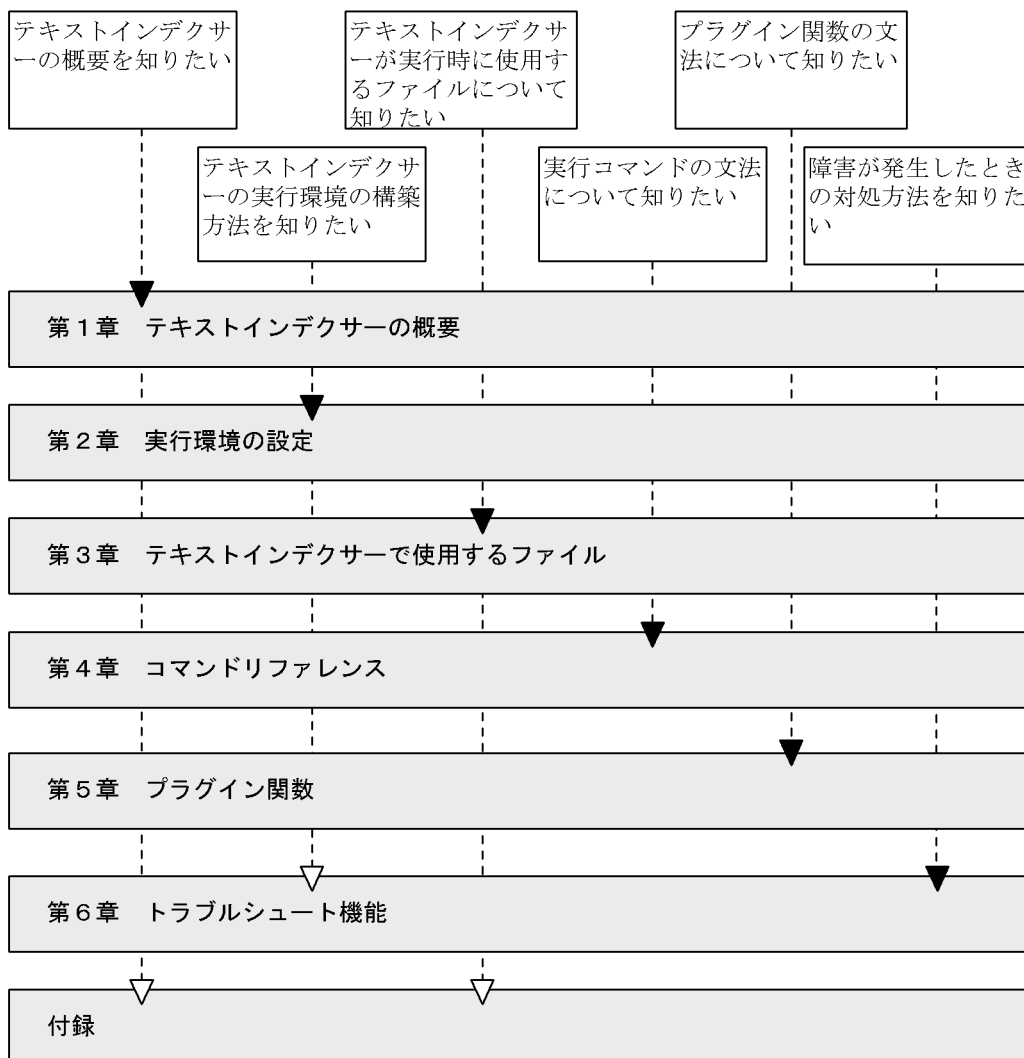
- Preprocessing Library for Text Search Version 2 (3000-7-270) ⁶

注 6

このマニュアルでは、「Preprocessing Library for Text Search」と表記しています。

読書手順

このマニュアルは、利用目的に合わせて章を選択してお読みいただけます。利用目的別に、次の流れに従ってお読みいただくことをお勧めします。



(凡例)



: 必ず読む項目



: 必要に応じて読む項目

このマニュアルでの表記

このマニュアルでは、製品名称を次に示す略称で表記しています。

製品名称	略称
uCosminexus DocumentBroker Development Kit Version 3	DocumentBroker

製品名称	略称
uCosminexus DocumentBroker Runtime Version 3	
uCosminexus DocumentBroker Server Version 3	
uCosminexus DocumentBroker Text Search Index Loader Version 3	
uCosminexus DocumentBroker Server Version 3	DocumentBroker サーバ
Document Filter for Text Search Version 3	Document Filter for Text Search
HiRDB/Single Server Version 6	HiRDB
HiRDB/Single Server Version 7	
HiRDB/Single Server Version 8	
HiRDB/Parallel Server Version 6	
HiRDB/Parallel Server Version 7	
HiRDB/Parallel Server Version 8	
HiRDB/Single Server - Object Option Version 6	HiRDB Object Option
HiRDB/Parallel Server - Object Option Version 6	
HiRDB/Single Server - Object Option Version 7	
HiRDB/Parallel Server - Object Option Version 7	
HiRDB/Run Time Version 6	HiRDB/Run Time
HiRDB/Run Time Version 7	
HiRDB/Run Time Version 8	
HiRDB Adapter for XML - Enterprise Edition	HiRDB Adapter for XML
HiRDB Adapter for XML - Standard Edition	
HiRDB Text Search Plug-in Version 2	HiRDB Text Search Plug-in
HiRDB Text Search Plug-in Version 7	
HiRDB Text Search Plug-in Version 8	
Microsoft(R) Excel	Excel
Microsoft(R) Office Excel	
Microsoft(R) Office PowerPoint(R)	PowerPoint
Microsoft(R) PowerPoint(R)	
Microsoft(R) Office Word	Word
Microsoft(R) Word	
Microsoft(R) Visual C++(R)	Visual C++
Microsoft(R) Windows Server(R) 2003, Enterprise Edition 日本語版	Windows Server 2003
Microsoft(R) Windows Server(R) 2003, Standard Edition 日本語版	

製品名称	略称
Microsoft(R) Windows Server(R) 2003 R2, Enterprise Edition 日本語版	Windows Server 2003 R2
Microsoft(R) Windows Server(R) 2003 R2, Standard Edition 日本語版	
Microsoft(R) Windows Server(R) 2003 R2, Enterprise x64 Edition 日本語版	Windows Server 2003 R2 または Windows Server 2003 R2 (x64)
Microsoft(R) Windows Server(R) 2003 R2, Standard x64 Edition 日本語版	
Microsoft(R) Windows Server(R) 2008 Enterprise 32-bit 日本語版	Windows Server 2008 または Windows Server 2008 x86
Microsoft(R) Windows Server(R) 2008 Standard 32-bit 日本語版	
Preprocessing Library for Text Search Version 2	Preprocessing Library for Text Search

なお、このマニュアルでは、Windows Server 2003、Windows Server 2003 R2、Windows Server 2003 R2 (x64)、および Windows Server 2008 x86 を合わせて Windows と表記します。

uCosminexus DocumentBroker のマニュアルで使用する略語

このマニュアルで使用する英略語を次に示します。

英略語	英字での表記
ACE	<u>A</u> ccess <u>C</u> ontrol <u>E</u> lement
ACFlag	<u>A</u> ccess <u>C</u> ontrol <u>F</u> lag
ACL	<u>A</u> ccess <u>C</u> ontrol <u>L</u> ist
AIIM	<u>A</u> ssociation for <u>I</u> nformation and <u>I</u> mage <u>M</u> anagement International
API	<u>A</u> pplication <u>P</u> rogramming <u>I</u> nterface
ASCII	<u>A</u> merican <u>S</u> tandard <u>C</u> ode for <u>I</u> nformation <u>I</u> nterchange
BES	<u>B</u> ack <u>E</u> nd <u>S</u> erver
BLOB	<u>B</u> inary <u>L</u> arge <u>O</u> bject
BNF	<u>B</u> ackus <u>N</u> ormal <u>F</u> orm
BOA	<u>B</u> asic <u>O</u> bject <u>A</u> dapter
CD-ROM	<u>C</u> ompact <u>D</u> isc <u>R</u> ead <u>O</u> nly <u>M</u> emory
CGI	<u>C</u> ommon <u>G</u> ateway <u>I</u> nterface
CORBA	<u>C</u> ommon <u>O</u> bject <u>R</u> equest <u>B</u> roker <u>A</u> rchitecture
CPU	<u>C</u> entral <u>P</u> rocessing <u>U</u> nit
CR	<u>C</u> arriage <u>R</u> eturn

英略語	英字での表記
CSV	<u>C</u> omma <u>S</u> eparated <u>V</u> alue
DAP	<u>D</u> irectory <u>A</u> ccess <u>P</u> rotocol
DB	<u>D</u> atabase
DBMS	<u>D</u> atabase <u>M</u> anagement <u>S</u> ystem
DCD	<u>D</u> ocument <u>C</u> ontent <u>D</u> escription
DIT	<u>D</u> irectory <u>I</u> nformation <u>T</u> ree
DLL	<u>D</u> ynamic <u>L</u> inking <u>L</u> ibrary
DMA	<u>D</u> ocument <u>M</u> anagement <u>A</u> lliance
DN	<u>D</u> istinguished <u>N</u> ame
EOF	<u>E</u> nd of <u>F</u> ile
ESIS-B	<u>E</u> lement <u>S</u> tructure <u>I</u> nformation <u>S</u> et- <u>B</u> inary Format
GIF	<u>G</u> raphics <u>I</u> nterchange <u>F</u> ormat
GUI	<u>G</u> raphical <u>U</u> ser <u>I</u> nterface
GUID	<u>G</u> lobally <u>U</u> nique <u>I</u> dentifier
HTML	<u>H</u> ypertext <u>M</u> arkup <u>L</u> anguage
HTTP	<u>H</u> ypertext <u>T</u> ransfer <u>P</u> rotocol
IANA	<u>I</u> nternet <u>A</u> ssigned <u>N</u> umbers <u>A</u> uthority
ID	<u>I</u> dentifier
ISO	<u>I</u> nternational <u>O</u> rganization for <u>S</u> tandardization
JIS	<u>J</u> apanese <u>I</u> ndustrial <u>S</u> tandards
JPEG	<u>J</u> oint <u>P</u> hotographic <u>E</u> xpert <u>G</u> roup
LAN	<u>L</u> ocal <u>A</u> rea <u>N</u> etwork
LDAP	<u>L</u> ightweight <u>D</u> irectory <u>A</u> ccess <u>P</u> rotocol
LF	<u>L</u> ine <u>F</u> eed
MFC	<u>M</u> icrosoft <u>F</u> oundation <u>C</u> lass
MIME	<u>M</u> ultipurpose <u>I</u> nternet <u>M</u> ail <u>E</u> xtensions
OCR	<u>O</u> ptical <u>C</u> haracter <u>R</u> eaders
OIID	<u>O</u> bject <u>I</u> nstance <u>I</u> dentifier
OLE	<u>O</u> bject <u>L</u> inking and <u>E</u> MBEDDING
OMG	<u>O</u> bject <u>M</u> anagement <u>G</u> roup
ORB	<u>O</u> bject <u>R</u> equest <u>B</u> roker
ORDB	<u>O</u> bject <u>R</u> elational <u>D</u> atabase
OS	<u>O</u> perating <u>S</u> ystem

英略語	英字での表記
OTS	<u>O</u> bject <u>T</u> ransaction <u>S</u> ervice
PC	<u>P</u> ersonal <u>C</u> omputer
PDF	<u>P</u> ortable <u>D</u> ocument <u>F</u> ormat
RDB	<u>R</u> elational <u>D</u> atabase
RDN	<u>R</u> elative <u>D</u> istinguished <u>N</u> ame
RFC	<u>R</u> equest <u>f</u> or <u>C</u> omment
RTF	<u>R</u> ich <u>T</u> ext <u>F</u> ormat
SGML	<u>S</u> tandard <u>G</u> eneralized <u>M</u> arkup <u>L</u> anguage
SMTP	<u>S</u> imple <u>M</u> ail <u>T</u> ransfer <u>P</u> rotocol
SQL	<u>S</u> tructured <u>Q</u> uery <u>L</u> anguage
TCP/IP	<u>T</u> ransmission <u>C</u> ontrol <u>P</u> rotocol/ <u>I</u> nternet <u>P</u> rotocol
TIFF	<u>T</u> ag <u>I</u> mage <u>F</u> ile <u>F</u> ormat
UCS-2	<u>U</u> niversal <u>C</u> haracter <u>S</u> et coded in <u>2</u> octets
UCS-4	<u>U</u> niversal <u>C</u> haracter <u>S</u> et coded in <u>4</u> octets
UOC	<u>U</u> ser <u>O</u> wn <u>C</u> oding
URL	<u>U</u> niform <u>R</u> esource <u>L</u> ocator
UTC	<u>U</u> niversal <u>T</u> ime <u>C</u> oordinated
UTF-8	<u>8</u> -bit <u>U</u> CS <u>T</u> ransformation <u>F</u> ormat
W3C	<u>W</u> orld <u>W</u> ide <u>W</u> eb <u>C</u> onsortium
WWW	<u>W</u> orld <u>W</u> ide <u>W</u> eb
XML	<u>E</u> xtensible <u>M</u> arkup <u>L</u> anguage

このマニュアルで使用する記号

このマニュアルで使用する記号を次に示します。

記号	意味
	横に並べられた複数の項目に対する項目間の区切りを示し、「または」を意味します。 (例) A B A または B を指定することを示します。
{ }	この記号で囲まれている複数の項目のうちから一つを選択することを示します。項目が横に並べられ、記号 で区切られている場合は、そのうちの一つを選択します。 (例){ A B C } A, B または C のどれかを指定することを示します。

記号	意味
{ }	この記号で囲まれている項目は省略してもよいことを意味します。複数の項目が横に並べて記述されている場合には、すべてを省略するか、記号{ }と同じくどれか一つを選択します。 (例1){A} 「何も指定しない」が「Aを指定する」ことを示します。 (例2){B C} 「何も指定しない」が「BまたはCを指定する」ことを示します。
—	括弧で囲まれた複数項目のうち1項目に対し使用され、括弧内のすべてを省略したときシステムが取る標準値を示します。 (例){ <u>A</u> B} 「何も指定しない」が「AまたはBを指定する」ことを示します。何も指定しない場合はAが仮定されます。
: :=	この記号の左にあるものを右にあるもので定義することを示します。 (例) A := B 「AとはBである」と定義することを示します。
...	記述が省略されていることを示します。 (例) ABC... ABCの後ろに記述があり、その記述が省略されていることを示します。

このマニュアルで使用する構文要素

このマニュアルで使用する構文要素の種類を次に示します。

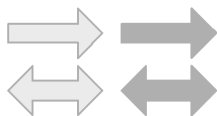
種類	定義
英 字	A ~ Z a ~ z
英小文字	a ~ z
英大文字	A ~ Z
数 字	0 ~ 9
英数字	A ~ Z a ~ z 0 ~ 9
記 号	! " # \$ % & ' () + , - . / : ; < = > @ [] ^ _ { } 空白

注 かたかな文字の直後にある「ー」は、かたかな文字になります。

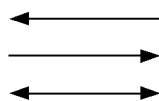
このマニュアルの図中で使用する記号

このマニュアルで使用する記号を、次のように定義します。

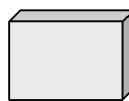
● データの流れ



● その他の流れ



● プログラム



● データベース



常用漢字以外の漢字の使用について

このマニュアルでは、常用漢字を使用することを基本としていますが、次に示す用語については、常用漢字以外の漢字を使用しています。

個所（かしよ）

KB（キロバイト）などの単位表記について

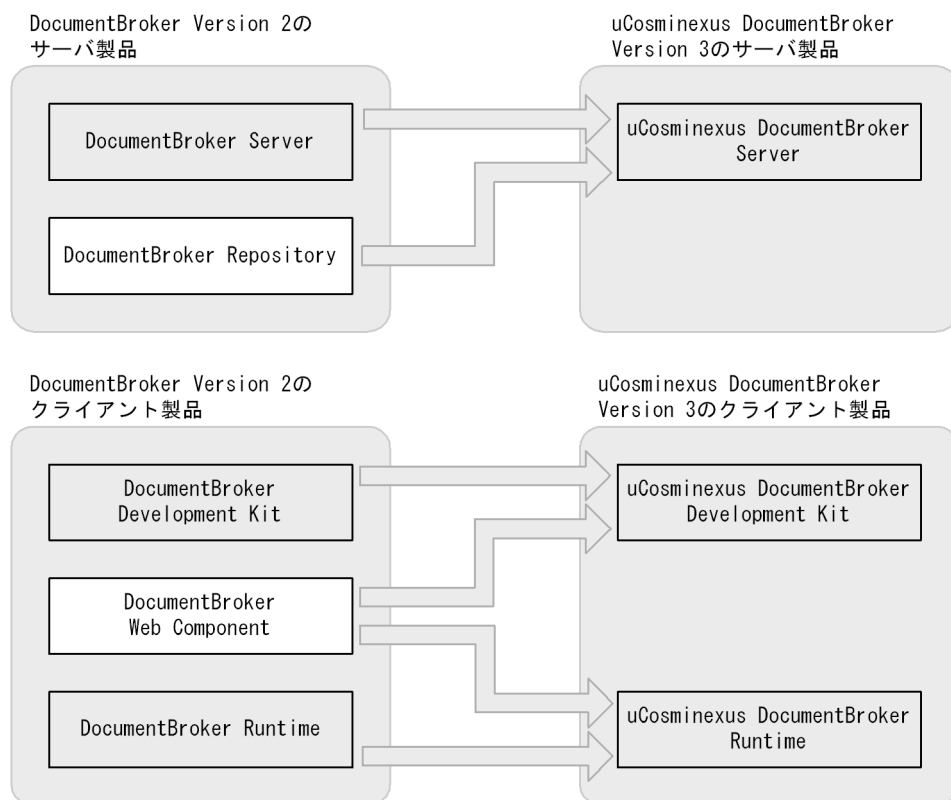
1KB（キロバイト）、1MB（メガバイト）、1GB（ギガバイト）、1TB（テラバイト）はそれぞれ 1,024 バイト、1,024² バイト、1,024³ バイト、1,024⁴ バイトです。

DocumentBroker Version 2 と uCosminexus DocumentBroker Version 3 の製品体系の違い

uCosminexus DocumentBroker Version 3 では次のように製品の統合を行いました。

- DocumentBroker Repository を uCosminexus DocumentBroker Server に統合しました。
- DocumentBroker Web Component を uCosminexus DocumentBroker Development Kit および uCosminexus DocumentBroker Runtime に統合しました。

DocumentBroker Version 2 と uCosminexus DocumentBroker Version 3 の製品体系の違いを次に示します。



DocumentBroker Version 2 と uCosminexus DocumentBroker Version 3 のマニュアルの対応

バージョンアップおよび製品体系の変更に伴い、uCosminexus DocumentBroker Version 3 では次に示すようにマニュアル名称を変更しました。

Version 2 のマニュアル名称	Version 3 のマニュアル名称
DocumentBroker Version 2 システム導入・運用ガイド	uCosminexus DocumentBroker Version 3 システム導入・運用ガイド
DocumentBroker Version 2 クラスライブラリ解説	uCosminexus DocumentBroker Version 3 クラスライブラリ C++ 解説
DocumentBroker Version 2 クラスライブラリリファレンス 基本機能編	uCosminexus DocumentBroker Version 3 クラスライブラリ C++ リファレンス 基本機能編
DocumentBroker Version 2 クラスライブラリリファレンス 概念 SGML 文書管理機能編	廃版
DocumentBroker Version 2 オブジェクト操作ツール	uCosminexus DocumentBroker Version 3 オブジェクト操作ツール
DocumentBroker Version 2 統計解析ツール	uCosminexus DocumentBroker Version 3 統計解析ツール
DocumentBroker Version 2 メッセージ	uCosminexus DocumentBroker Version 3 メッセージ
DocumentBroker Web Component Version 2 解説	uCosminexus DocumentBroker Version 3 クラスライブラリ Java 解説
DocumentBroker Web Component Version 2 リファレンス	uCosminexus DocumentBroker Version 3 クラスライブラリ Java リファレンス
DocumentBroker Web Component Version 2 サンプル Web アプリケーション	uCosminexus DocumentBroker Version 3 クラスライブラリ Java サンプル Web アプリケーション
DocumentBroker Text Search Index Loader Version 2	uCosminexus DocumentBroker Text Search Index Loader Version 3
DocumentBroker Rendering Option システム導入・運用ガイド	DocumentBroker Rendering Option Version 3
	DocumentBroker Rendering Option Version 3 (PDFstaff 編)
DocumentBroker Object Loader Version 2	uCosminexus DocumentBroker Object Loader Version 3

印の製品については、サポート時期をご確認ください。

目次

1	テキストインデクサーの概要	1
1.1	テキストインデクサーとは	2
1.2	システム構成	3
1.3	テキストインデクサーの機能	5
1.3.1	テキストインデクサーが提供するコマンド	5
1.3.2	フィルタプラグイン機能	5
1.4	テキストインデクサーの処理概要	6
1.4.1	処理概要	6
1.4.2	ファイル環境	7
1.5	テキストインデクサーで使用できる文字コード種別	9
2	実行環境の設定	11
2.1	実行環境設定の流れ	12
2.2	インストールとアンインストール	13
2.2.1	インストール	13
2.2.2	インストールディレクトリと提供ファイルの内容	14
2.2.3	アンインストール	16
2.3	DocumentBroker サーバの設定	18
2.4	テキストインデクサーの環境変数の設定	19
2.5	HiRDB の環境の設定	20
2.5.1	全文検索用の表の作成	20
2.5.2	クライアント側の環境設定	20
2.5.3	HiRDB Text Search Plug-in での環境設定	20
2.6	実行環境および運用に関する注意事項	21
3	テキストインデクサーで使用するファイル	29
3.1	使用するファイルの一覧	30
3.2	ファイル形式と文法	31
3.2.1	ファイル形式	31
3.2.2	文法	31
3.3	環境設定ファイル	33
3.3.1	記述形式	33

3.3.2	記述規則	33
3.3.3	環境定義パラメタ	33
3.3.4	抽出処理が失敗した文書の取り扱い	39
3.3.5	環境設定ファイルの記述例	39
3.4	RenditionType 定義ファイル	41
3.4.1	提供するフィルタ拡張子と使用するテキスト抽出ライブラリ	41
3.4.2	記述形式	41
3.4.3	記述規則	42
3.4.4	RenditionType 定義ファイルの記述例	43
3.5	定義ファイル	44
3.5.1	記述形式	44
3.5.2	記述規則	44
3.5.3	定義ファイルの記述例	45
3.6	処理対象文書クラス定義ファイル	46
3.6.1	記述形式	46
3.6.2	記述規則	46
3.6.3	処理対象文書クラス定義ファイルの記述例	46
3.7	環境変数ファイル	48
3.7.1	記述形式	48
3.7.2	記述規則	48
3.7.3	環境変数ファイルに設定するパラメタ	49
3.8	プラグイン環境設定ファイル	52
3.8.1	記述形式	52
3.8.2	記述規則	52
3.8.3	プラグイン環境設定ファイルの記述例	55

4	コマンドリファレンス	57
	実行コマンド一覧	58
	コマンドの形式	59
	EDMTextIndex (テキストインデクサーの実行)	61
	EDMCrtTIDF (テキストインデクサー定義ファイル生成)	65

5	プラグイン関数	67
	テキストインデクサーが定義するプラグイン関数	68
	プラグイン関数の処理	70

説明する項目	72
フィルタプラグイン関数 (dbrIdxTextFilter())	73
プロパティ取得関数 (任意の名称)	76

6

トラブルシューティング機能	79
6.1 トラブルシューティング機能の概要	80
6.2 障害処理	81
6.3 メッセージの形式	82
6.3.1 メッセージの出力形式	82
6.3.2 メッセージの記述形式	82
6.4 メッセージの詳細	84

付録

付録 A 処理対象の全文検索インデクス用プロパティ	133
付録 B 文書空間で使用する文字コード種別が UTF-8 の場合のサポート範囲	134
付録 C 用語解説	135

索引

索引	139
----	-----

目次

図 1-1	テキストインデクサーのシステム構成	3
図 1-2	テキストインデクサーの処理概要	7
図 2-1	実行環境設定の流れ	12
図 2-2	テキストインデクサーのインストールディレクトリ	15
図 2-3	一つのセル内が複数行で構成された表の例	25
図 2-4	文書オブジェクトの構成例	26
図 3-1	環境設定ファイルの記述形式	33
図 3-2	書誌情報の作成例	39
図 3-3	RenditionType 定義ファイルの記述形式	42
図 3-4	定義ファイルの記述形式	44
図 3-5	定義ファイルの処理対象文書クラスの定義方法	45
図 3-6	定義ファイルの書誌情報プロパティの定義方法	45
図 3-7	処理対象文書クラス定義ファイルの記述形式	46
図 3-8	環境変数ファイル (NTconfig.ini) の記述形式	48
図 3-9	プラグイン環境設定ファイル (Pluginconfig.ini) の記述形式	52

表目次

表 1-1	テキストインデクサーの前提 PP 構成一覧	4
表 1-2	テキストインデクサーが提供するコマンドの一覧	5
表 3-1	テキストインデクサーで使用するファイル	30
表 3-2	提供するフィルタ拡張子と使用するテキスト抽出ライブラリ	41
表 3-3	使用できるワイルドカード	42
表 4-1	テキストインデクサーが提供するコマンドの一覧	58
表 4-2	終了コード一覧	60
表 4-3	指定できるオプションと省略時の処理	62
表 6-1	要因コード別の対策一覧	85
表 6-2	要因コード別の対策一覧	88
表 6-3	要因コード別の対策一覧	89
表 6-4	要因コード別の対策一覧	90
表 6-5	要因コード別の対策一覧	92
表 6-6	要因別の対策一覧	97
表 6-7	要因別の対策一覧	97
表 6-8	要因コード別の対策一覧	98
表 6-9	要因コード別の対策一覧	99
表 6-10	エラー情報別の対策一覧	101
表 6-11	%1 に示されるエラーメッセージ	104
表 6-12	要因コード別の対策一覧	107
表 6-13	要因コード別の対策一覧	112
表 6-14	要因別の対策一覧	116
表 6-15	要因別の対策一覧	117
表 6-16	要因別の対策一覧	119
表 6-17	要因別の対策一覧	119
表 6-18	要因別の対策一覧	120
表 6-19	要因コード別の対策一覧	121
表 6-20	要因別の対策一覧	122
表 6-21	要因別の対策一覧	123
表 6-22	要因別の対策一覧	124
表 6-23	要因別の対策一覧	125
表 6-24	要因コード別の対策一覧	126
表 6-25	要因コード別の対策一覧	131

表 A-1	テキストインデクサーが処理対象とする全文検索インデクス用プロパティ	134
表 B-1	文書空間で使用する文字コード種別が UTF-8 の場合に使用できるテキスト抽出ライブラリ	135

1

テキストインデクサーの概要

この章では、テキストインデクサーの機能と処理概要を説明します。

1.1 テキストインデクサーとは

1.2 システム構成

1.3 テキストインデクサーの機能

1.4 テキストインデクサーの処理概要

1.5 テキストインデクサーで使用できる文字コード種別

1.1 テキストインデクサーとは

テキストインデクサーは DocumentBroker サーバに格納されている文書を全文検索インデクスに登録するユティリティプログラムです。

テキストインデクサーは、種々のアプリケーションで作成された文書からテキスト情報を抽出して全文検索用のインデクスファイルを作成し、これを DocumentBroker サーバの全文検索インデクスに登録する機能を提供します。

この機能によって、ワードプロセッサなどのアプリケーションで作成した文書も、ユーザプログラムでテキスト情報を抽出して全文検索インデクスに登録することなく、DocumentBroker サーバに登録できるようになります。

テキストインデクサーでインデクス登録できるドキュメントの種類には、プレーンテキスト文書や XML 文書、WORD 文書などがあります。これらテキストインデクサーが扱うドキュメントの種類は、指定するフィルタ拡張子とテキスト抽出ライブラリの組み合わせに従います。フィルタ拡張子とテキスト抽出ライブラリの組み合わせについては、「3.4.1 提供するフィルタ拡張子と使用するテキスト抽出ライブラリ」を参照してください。なお、フィルタプラグイン機能を利用すると、ユーザが作成するフィルタプラグイン関数でテキスト情報を抽出できるドキュメントの種類が処理対象となります。フィルタプラグイン機能の詳細については、「1.3.2 フィルタプラグイン機能」を参照してください。

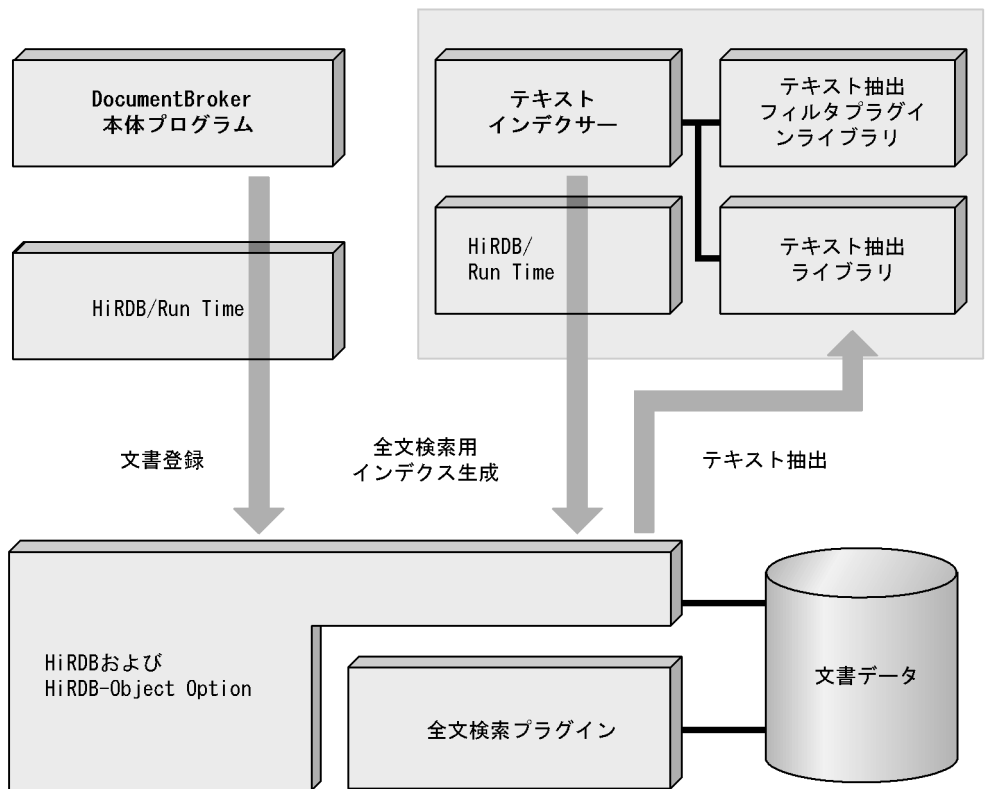
注 Windows Server 2008 を使用する場合は、XML 文書のインデクス登録はできません。

1.2 システム構成

テキストインデクサーは、全文検索インデクスをデータベースに登録するために、文書データのテキストを抽出する機能（ライブラリ）であるテキスト抽出ライブラリを使用します。

テキストインデクサーが動作するためのシステム構成を図 1-1 に示します。また、表 1-1 に前提となる PP（プログラムプロダクト）の構成を示します。

図 1-1 テキストインデクサーのシステム構成



1. テキストインデクサーの概要

表 1-1 テキストインデクサーの前提 PP 構成一覧

前提機能	対応する製品名称
HiRDB	HiRDB/Single Server Version 6 , HiRDB/Parallel Server Version 6 , HiRDB/Single Server Version 7 , HiRDB/Parallel Server Version 7 , HiRDB/Single Server Version 8 , HiRDB/Parallel Server Version 8 , HiRDB/Run Time Version 6 , HiRDB/Run Time Version 7 , または HiRDB/Run Time Version 8
HiRDB - Object Option	HiRDB/Single Server - Object Option Version 6 , HiRDB/Parallel Server - Object Option Version 6 , HiRDB/Single Server - Object Option Version 7 , または HiRDB/Parallel Server - Object Option Version 7
全文検索プラグイン	HiRDB Text Search Plug-in Version 2 , HiRDB Text Search Plug-in Version 7 , または HiRDB Text Search Plug-in Version 8
DocumentBroker 本体プログラム	uCosminexus DocumentBroker Server Version 3
テキスト抽出ライブラリ	Document Filter for Text Search Version 3 Preprocessing Library for Text Search Version 2 HiRDB Adapter for XML - Enterprise Edition または HiRDB Adapter for XML - Standard Edition

テキストインデクサーの動作環境に関する注意

- DocumentBroker サーバの DocumentSpace 構成定義ファイル (docspase.ini) に記述した接続情報の定義を使用してデータベースに接続する場合、テキストインデクサーは DocumentBroker サーバと同一のマシン上で動作することが前提となります。
- テキストインデクサーの環境設定ファイル (TextIndex.ini) に記述した接続情報の定義を使用してデータベースに接続する場合、テキストインデクサーは DocumentBroker サーバとは別マシンでも動作できます。

1.3 テキストインデクサーの機能

この節では、テキストインデクサーの提供するコマンドの機能およびフィルタプラグイン機能の概要を説明します。

1.3.1 テキストインデクサーが提供するコマンド

テキストインデクサーが提供するコマンドの一覧を表 1-2 に示します。なお、コマンド仕様の詳細については、「4. コマンドリファレンス」を参照してください。

表 1-2 テキストインデクサーが提供するコマンドの一覧

項番	コマンド名	機能概要
1	EDMTextIndex	テキストインデクサーを実行する。
2	EDMCrtTIDF	定義ファイルを生成する。

1.3.2 フィルタプラグイン機能

テキストインデクサーが定義する関数に従い、ユーザが作成したプラグインライブラリを使用してテキスト抽出する機能を、フィルタプラグイン機能といいます。この機能は、次の場合に利用してください。

テキストインデクサーが対象としているドキュメント以外からテキスト抽出したい場合

テキストインデクサーが前提としていないテキスト抽出フィルタを使用したい場合

テキスト抽出後に文書オブジェクトのユーザプロパティ値を使用して、処理を実行したい場合

なお、フィルタプラグイン機能を利用するには、プラグイン環境設定ファイルを作成する必要があります。プラグイン環境設定ファイルの定義内容については、「3.8 プラグイン環境設定ファイル」を参照してください。また、テキストインデクサーが定義する関数については、「5. プラグイン関数」を参照してください。

1.4 テキストインデクサーの処理概要

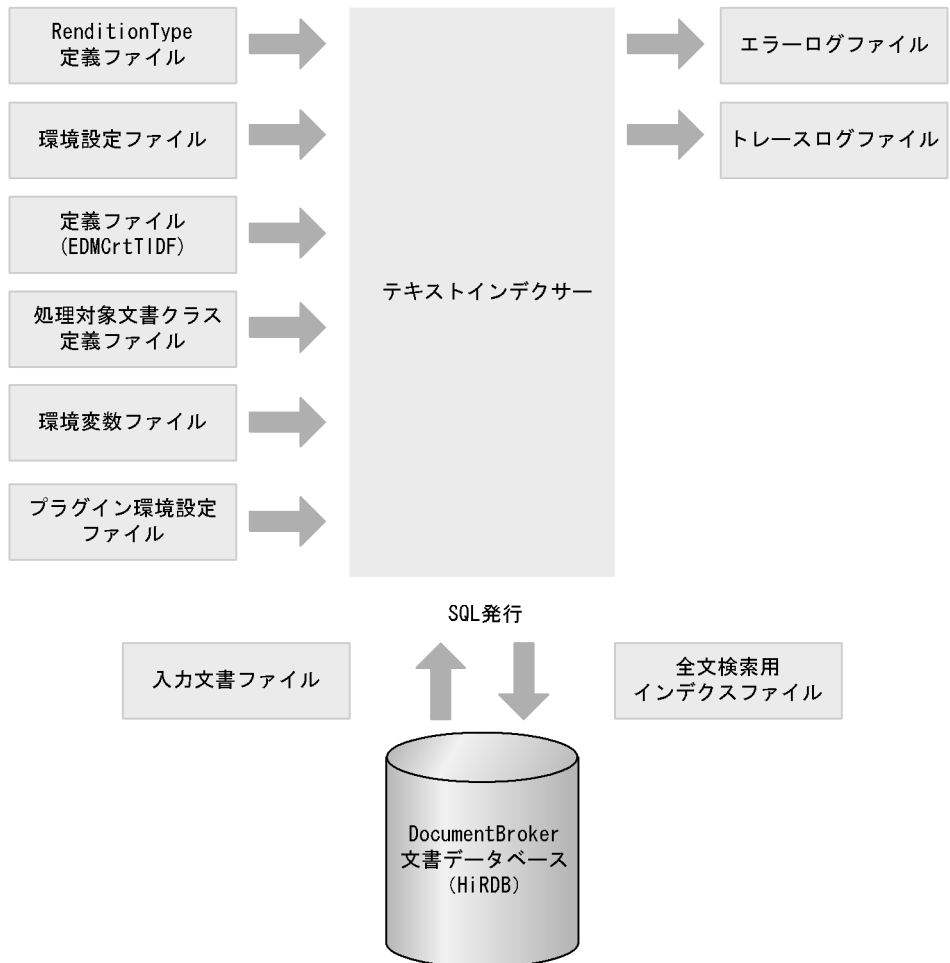
テキストインデクサーの処理概要およびテキストインデクサーを実行するために必要なファイルについて説明します。

1.4.1 処理概要

テキストインデクサーは RenditionType 定義ファイル，環境設定ファイル，定義ファイル，および処理対象文書クラス定義ファイルの情報に基づいて動作します。コマンドで指定された全文検索用の文書クラスから処理対象文書を取り出し，RenditionType 定義情報に基づいてコンテンツのテキストを抽出し，全文検索用インデクスファイルを作成してデータベースに登録します。実行時には，エラー情報やトレース情報も取得します。

テキストインデクサーの処理概要を図 1-2 に示します。

図 1-2 テキストインデクサーの処理概要



1.4.2 ファイル環境

テキストインデクサーを実行するために必要なファイルとエラー情報やトレース情報が取得されるファイルについて説明します。

(1) テキストインデクサーを実行するために必要なファイル

テキストインデクサーを実行するために必要なファイルについて説明します。

(a) RenditionType 定義ファイル

テキスト抽出ライブラリが処理できる RenditionType, またはテキストを示す RenditionType の一覧を定義するテキストファイルです。

1. テキストインデクサーの概要

(b) 環境設定ファイル

テキストインデクサー動作環境定義パラメタを定義するテキストファイルです。

(c) 定義ファイル

DocumentBroker サーバが提供するメタ情報ファイルに指定されたクラス名（プロパティ名）と GUID 値の対応を定義するテキストファイルです。EDMCRtTIDF コマンドを実行すると自動生成されます。このファイルに定義した処理対象文書クラス名および書誌情報プロパティ名を使用して、環境設定ファイル、処理対象文書クラス定義ファイル、およびプラグイン環境設定ファイルを作成します。

(d) 処理対象文書クラス定義ファイル

テキストインデクサーが処理対象とする全文検索対象クラスを、複数個実行する場合に作成するテキストファイルです。

(e) 入力文書ファイル

文書空間内に登録されている文書ファイルです。

(f) 全文検索用インデクスファイル

全文検索用の情報を格納したファイルです。HiRDB 登録後に削除される一時的なファイルです。

(g) 環境変数ファイル

テキストインデクサーが動作するために必要な情報を設定するファイルです。

(h) プラグイン環境設定ファイル

フィルタプラグイン機能を使用する場合に定義するテキストファイルです。

(2) エラー情報およびトレース情報が取得されるファイル

テキストインデクサーを実行したときに発生するエラー情報やトレース情報が取得されるファイルについて説明します。各ファイルの詳細については、「6. トラブルシュート機能」を参照してください。

(a) エラーログファイル

エラーログの取得を指定した場合、テキストインデクサー実行時に発生したメッセージはエラーログファイルに出力されます。

(b) トレースログファイル

テキストインデクサーのトレース情報を出力するファイルです。主に、テキストインデクサーの保守に使用します。

1.5 テキストインデクサーで使用できる文字コード種別

この節では、テキストインデクサーで使用できる文字コード種別について説明します。

テキストインデクサーで使用する文字コード種別は、DocumentBroker サーバの文書空間で使用する文字コード種別に従います。そのため、テキストインデクサーで文字コード種別を設定する必要はありません。

DocumentBroker サーバの文書空間で使用できる文字コード種別は次のどちらかです。

- Shift-JIS
- UTF-8 (使用できる文字コードの範囲は UCS-2 または UCS-4 です)

DocumentBroker サーバの文書空間で使用する文字コード種別の設定および運用方法については、マニュアル「DocumentBroker Version 3 システム導入・運用ガイド」を参照してください。

(1) 文書空間の文字コード種別に UTF-8 を設定した場合の注意事項

DocumentBroker サーバの文書空間の文字コード種別に UTF-8 を設定した場合の注意事項を次に示します。

- ユーザが記述する次の項目は、印刷可能な ASCII コードで記述してください。
 - コマンドに指定する引数
 - テキストインデクサーで使用する、次のファイルの記載内容
環境設定ファイル、RenditionType 定義ファイル、環境変数ファイル、定義ファイル、処理対象文書クラス定義ファイル、プラグイン環境設定ファイル
 - テキストインデクサーで使用する、次のファイルのファイル名
定義ファイル、処理対象文書クラス定義ファイル
- 文書のテキスト情報の抽出に使用できるテキスト抽出ライブラリに制限があります。テキスト抽出ライブラリのサポート範囲の詳細については、「付録 B 文書空間で使用する文字コード種別が UTF-8 の場合のサポート範囲」を参照してください。
- メッセージテキストの挿入句に文書ファイル名が表示される場合、文書空間の文字コード種別の指定内容によって文書ファイル名の表示形式が異なります。文書ファイル名の表示形式の詳細については、「6.3.2(3) メッセージテキストの説明」を参照してください。

2

実行環境の設定

この章では、テキストインデクサーを実行するために必要な環境設定について説明します。

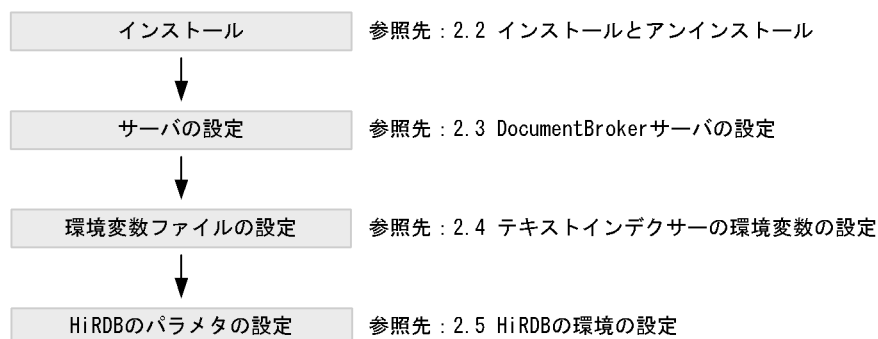
-
- 2.1 実行環境設定の流れ
 - 2.2 インストールとアンインストール
 - 2.3 DocumentBroker サーバの設定
 - 2.4 テキストインデクサーの環境変数の設定
 - 2.5 HiRDB の環境の設定
 - 2.6 実行環境および運用に関する注意事項
-

2.1 実行環境設定の流れ

この節では、テキストインデクサーの実行環境を設定するための流れについて説明します。

テキストインデクサーの実行環境を設定する順序を、図 2-1 に示します。

図 2-1 実行環境設定の流れ



説明：

1. インストールを実行したあと、環境を確認します。
インストール後に作成されるディレクトリと提供されるファイルを確認します。
2. DocumentBroker サーバの環境を設定します。
DocumentBroker サーバで必要な環境設定をします。
3. テキストインデクサーの環境変数ファイルを設定します。
環境変数ファイルの次のパラメタを設定します。
 - `_HIEDMS_TRACE_DIR`
 - `_HIEDMS_TRACE_NUM`
 - `_HIEDMS_TRACE_SIZE`
 - `_HIEDMS_TRACE_LEVEL`
 - `TMPDIR`
4. HiRDB および HiRDB Text Search Plug-in のパラメタを設定します。

2.2 インストールとアンインストール

この節では、テキストインデクサーのインストールとアンインストールについて説明します。

2.2.1 インストール

ここでは、テキストインデクサーのインストール方法について説明します。

< 操作 >

1. Administrators グループのユーザでログインします。
2. インストールを実行する前にすべての Windows アプリケーションを終了させてください。
3. インストール用の CD-ROM 内の Setup.exe を起動します。
4. インストールプログラムが起動したあと、画面の指示に従ってインストールしてください。
インストール説明のダイアログが表示されます。
5. 「次へ」ボタンをクリックします。
ユーザ情報として「名前」と「会社名」を入力するダイアログが表示されます。
上書きインストールの場合には、上書き確認ダイアログが表示されます。ここで「はい」ボタンをクリックした場合、「10. プログラムフォルダを指定します」の処理へスキップされます。
6. 「名前」と「会社名」を入力します。
7. 「次へ」ボタンをクリックします。
インストール先のディレクトリを指定するダイアログが表示されます。
8. インストール先ディレクトリを指定します。
指定したディレクトリの下にインストールされます。デフォルトのインストール先は C:\Program files\HITACHI\DocBroker_Tidx です。C:\ は、OS がインストールされているドライブ名です。
9. 「次へ」ボタンをクリックします。
プログラムフォルダを指定するダイアログが表示されます。
10. プログラムフォルダを指定します。
指定したプログラムフォルダにプログラムアイコンが追加されます。
11. 「次へ」ボタンをクリックします。
現在の設定（ユーザ情報、インストール先のディレクトリ、プログラムフォルダ）確認のダイアログが表示されます。
12. 設定を確認して、「次へ」ボタンをクリックします。
インストールが開始されます。

2. 実行環境の設定

インストールが終了すると、インストールが終了したことを通知するダイアログが表示されます。

- 13.「完了」ボタンをクリックします。
インストールプログラムが終了します。

- 14.Windows を再起動します。

注意事項

- DocumentBroker Text Search Index Loader は、次の製品と同じマシン上にインストールできません。
 - DocumentBroker Text Search Index Loader Version 2
 - ラビニティ ECM Index Loader
- すでに該当する製品がインストールされている場合、次のどれかの方法でインストールしてください。
 - 該当する製品をアンインストールしてからインストールする
 - 別のマシンにインストールする
- 上書きインストールの場合、インストールディレクトリの下の ¥etc, ¥tmp および ¥spool は上書きされません。
- バージョンアップによる上書きインストールを実施した場合、旧バージョンのマニフェストデータが残ります。このため、旧バージョンをアンインストールしてから新バージョンをインストールすることを推奨します。

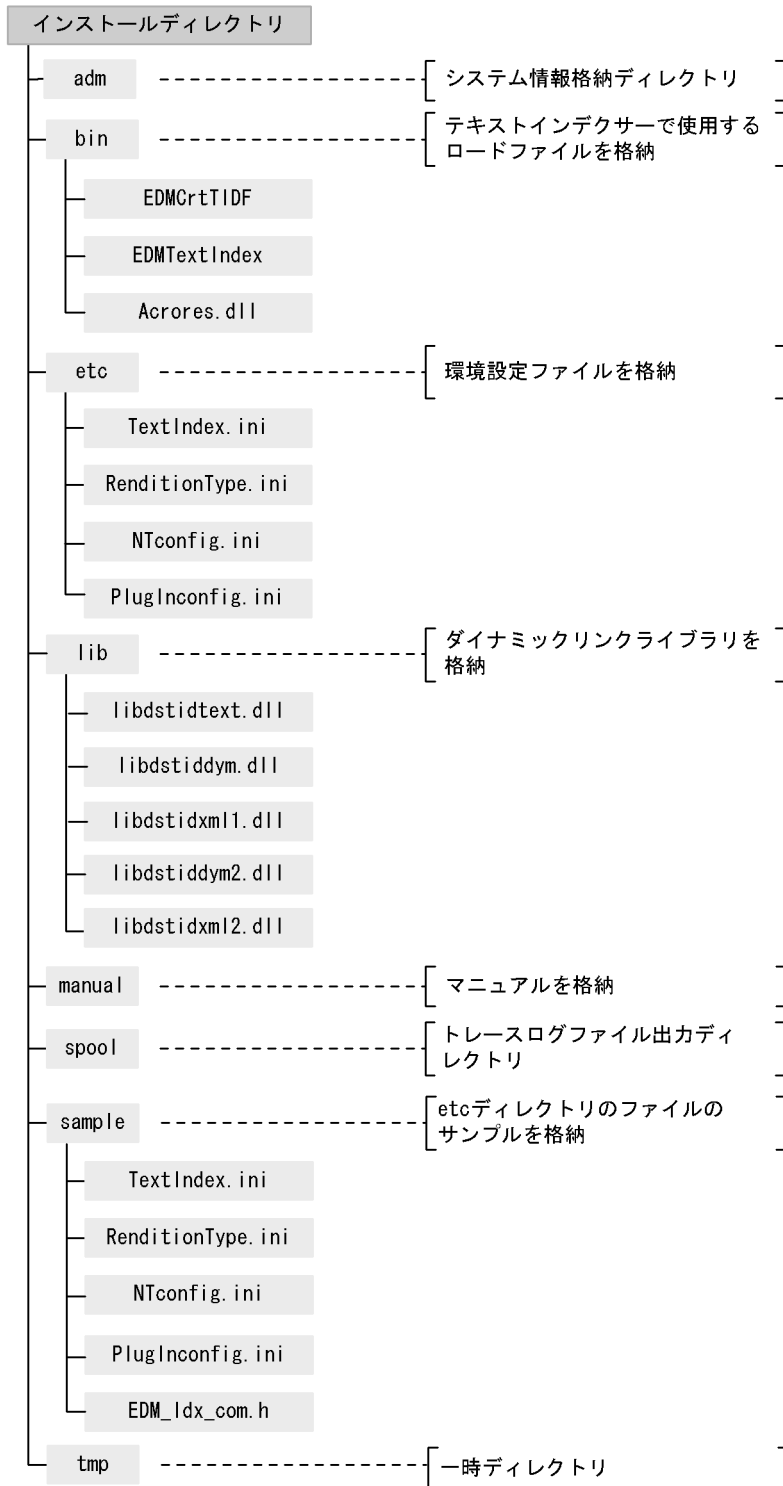
2.2.2 インストールディレクトリと提供ファイルの内容

(1) インストールディレクトリ

テキストインデクサーのインストールが終了したあとにディレクトリが正しく作成されているか確認します。

テキストインデクサーをインストールしたあとのインストールディレクトリ下の構成を図 2-2 に示します。

図 2-2 テキストインデクサーのインストールディレクトリ



2. 実行環境の設定

(2) 提供ファイル

テキストインデクサーが提供するファイルについて次に示します。

提供ファイルは、インストールディレクトリの `¥etc` の下、および `¥sample` の下に格納されます。実行環境に合わせて、`¥etc` の下にある各提供ファイルの設定内容を変更してください。

`¥sample` の下にある提供ファイルは直接変更しないでください。`¥sample` の下の提供ファイルは、変更した `¥etc` の下のファイルを初期状態（インストール直後の状態）に戻す場合に使用してください。

TextIndex.ini

テキストインデクサー制御用のパラメタを定義した環境設定ファイルです。環境に合わせて設定内容を変更してください。環境設定ファイル（TextIndex.ini）の設定については「3.3 環境設定ファイル」を参照してください。

RenditionType.ini

テキスト抽出ライブラリがサポートしているアプリケーションに対応する RenditionType リストのサンプルファイルです。指定したレンディションに応じて定義してください。

NTconfig.ini

テキストインデクサーの動作に必要な環境変数を設定するための環境変数ファイルです。動作環境に合わせて、設定内容を変更してください。環境変数ファイル（NTconfig.ini）の設定については「3.7 環境変数ファイル」を参照してください。

PlugInconfig.ini

フィルタプラグイン機能を使用する場合に定義するテキストファイルです。プラグイン環境設定ファイル（PlugInconfig.ini）の設定については、「3.8 プラグイン環境設定ファイル」を参照してください。

2.2.3 アンインストール

ここでは、テキストインデクサーのアンインストール方法について説明します。

<操作>

1. Administrators グループのユーザでログインします。
2. 「コントロールパネル」の「プログラムの追加と削除」を実行します。
アプリケーションの追加と削除のプロパティが表示されます。
3. インストールと削除タブでアンインストールするプログラム（DocumentBroker Text Search Index Loader Version 3）を選択して「追加と削除」ボタンをクリックします。

注意事項

アンインストール時にインストールディレクトリおよびその下にある `¥etc`、`¥spool`、

¥tmp は、自動的に削除されません。必要に応じて、ユーザがこれらのディレクトリを削除してください。

2.3 DocumentBroker サーバの設定

この節では、DocumentBroker サーバの設定について説明します。

(1) 環境変数の設定

(a) 環境変数の設定場所

Windows Server 2003, Windows Server 2003 R2 の場合

環境変数は、「コントロールパネル」にある「システム」の詳細設定タブの「環境変数」ボタンをクリックして表示される「環境変数」ダイアログに設定されています。

Windows 2008 の場合

環境変数は、「コントロールパネル」にある「システム」 - 「システムの詳細設定」の詳細設定タブの「環境変数」ボタンをクリックして表示される「環境変数」ダイアログに設定されています。

(b) 確認する環境変数

DocumentBroker サーバがインストール済みの場合は、次の環境変数が設定されているか確認してください。

- DOCBROKERDIR

DocumentBroker サーバの接続情報の定義を使用してデータベースに接続する場合は、DocumentBroker サーバのインストールディレクトリが設定されていることを確認してください。

テキストインデクサーの接続情報の定義を使用してデータベースに接続する場合は、この設定は必要ありません。

(2) 全文検索インデクスを登録する場合の認可識別子

DocumentBroker サーバの接続情報の定義を使用してデータベースに接続する場合

「<インストールディレクトリ> \etc\docspace.ini」に指定された認可識別子を使用します。そのため、HiRDB Text Search Plug-in を登録したユーザ ID を、docspace.ini の PdTSPluginOwner に指定してください。

テキストインデクサーの接続情報の定義を使用してデータベースに接続する場合

「<インストールディレクトリ> \etc\TextIndex.ini」に指定された認可識別子を使用します。そのため、HiRDB Text Search Plug-in を登録したユーザ ID を、TextIndex.ini の PdTSPluginOwner に指定してください。

2.4 テキストインデクサーの環境変数の設定

この節では、テキストインデクサーを使用するための環境変数の設定について説明します。

必要に応じて環境変数ファイル (NTconfig.ini) に次のパラメータを設定してください。設定内容については「3.7 環境変数ファイル」を参照してください。

なお、インストール直後の環境変数ファイルはコメント行になっています。

_HIEDMS_TRACE_DIR

トレースログファイルの出力先ディレクトリを設定します。

_HIEDMS_TRACE_NUM

トレースの出力情報がトレースファイルのサイズの上限を超えた場合、トレースの出力先を別のファイルに切り替えられます。この場合の切り替えるファイル数を設定します。

_HIEDMS_TRACE_SIZE

トレースを出力するファイルのサイズを設定します。

_HIEDMS_TRACE_LEVEL

トレースの出力レベルを設定します。

TMPDIR

次の用途として設定します。

- PDF 文書または XML 文書のテキスト抽出用の一時ディレクトリ
- フィルタプラグイン機能でテキストインデクサーとテキスト抽出フィルタプラグインライブラリとのデータの受け渡しをファイルで行う指定の場合の一時ディレクトリ

2.5 HiRDB の環境の設定

この節では、HiRDB の環境設定について説明します。

2.5.1 全文検索用の表の作成

全文検索用の表を作成する場合は、文書登録種別に「PLAIN」または「SGML」を指定してください。

テキストインデクサーが処理対象とする全文検索インデクス用のプロパティについては、「付録 A 処理対象の全文検索インデクス用プロパティ」を参照してください。また、表の作成方法については、マニュアル「HiRDB Text Search Plug-in」を参照してください。

2.5.2 クライアント側の環境設定

HiRDB クライアントで、次の環境変数を設定してください。

- PDSWAITTIME

PDSWAITTIME に 1 文書の処理時間よりも小さい値が設定されていると DB サーバとの接続が遮断される場合があります。このため、PDSWAITTIME にあらかじめ十分な値を設定しておくことを推奨します。詳細については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

2.5.3 HiRDB Text Search Plug-in での環境設定

テキストインデクサーでは、全文検索インデクスをデータベースに登録するために、HiRDB のプラグインプログラムである HiRDB Text Search Plug-in が必要です。

HiRDB Text Search Plug-in の環境設定については、マニュアル「HiRDB Text Search Plug-in」を参照してください。

2.6 実行環境および運用に関する注意事項

この節では、テキストインデクサーの実行環境および運用に関する注意事項について説明します。

(1) テキストインデクサーの実行者

テキストインデクサーの実行者は、Administrators グループのメンバーである必要があります。

(2) テキストインデクサーの実行環境を作成するディレクトリについて

インストールディレクトリがテキストインデクサーの実行環境を作成するディレクトリになります。

(3) HiRDB の起動

テキストインデクサー実行時は、HiRDB を起動しておく必要があります。

(4) テキストインデクサー (EDMTextIndex) の並列実行

テキストインデクサー (EDMTextIndex) は、並列実行できません。同時に複数のテキストインデクサーを実行した場合はエラーになります。

(5) テキストインデクサー (EDMTextIndex) のテキスト抽出処理

テキストインデクサー (EDMTextIndex) は、テキスト抽出処理に時間が掛かることがあるので DocumentBroker サーバの運用停止中にバッチジョブで実行することを推奨します。

(6) 効率的なインデクス登録処理について

インデクス登録処理では、テキストインデクサー (EDMTextIndex) 実行時に遅延一括作成機能を使用することで、効率的にインデクス登録できます。遅延一括作成機能を使用する場合は、HiRDB サーバ定義で `pd_plugin_ixmk_dir` を定義し、クライアントの環境変数 `PDPLGIXMK=YES` を設定した環境でテキストインデクサーを実行します。テキストインデクサー終了後に、HiRDB データベース再編成ユティリティを使用して遅延一括作成を実行します。

ただし、テキストインデクサーでインデクスファイルを作成し、現在遅延一括作成待ちとなっているインデクスファイル領域に対して、ユーザが別途書き込み処理をしようとすると HiRDB の遅延一括作成機能の仕様によってエラーになります。また、RD エリア単位に排他制御されるため、インデクスファイルが更新されているときは、そのインデクスファイルは検索対象とならなくなることがあります。

HiRDB データベース再編成ユティリティについては、マニュアル「HiRDB コマンドリファレンス」を参照してください。

2. 実行環境の設定

(7) テキストインデクサーのコマンド実行時間

テキストインデクサーの実行時に全体の処理時間の上限値を設定できます。ただし、テキストインデクサー実行中に、指定した処理時間の上限値に達すると処理を打ち切ります。それまでに処理した文書については全文検索インデクスへの登録を実行します。

(8) MaxOfLength とテキスト抽出の結果について

テキスト抽出の結果、テキスト抽出ライブラリの付加情報と抽出したテキストとの合計の長さが、MaxOfLength の値を超えない場合、抽出結果のすべてが格納されます。このとき格納されるテキストは、抽出されたテキストより付加情報分、テキスト長が長くなります。

テキスト抽出の結果、テキスト抽出ライブラリの付加情報と抽出したテキストとの合計の長さが、MaxOfLength の値を超える場合、MaxOfLength の値に合わせてテキスト抽出が打ち切られます。

ただし、XML 文書のテキスト抽出では、メッセージ KMBV12009-I 出力後に KMBV12037-W が出力され、テキスト抽出結果は登録されません。この場合は、MaxOfLength の値を KMBV12037-W のテキストデータ長に表示された値以上に設定して再度実行してください。

(9) テキストインデクサーを実行する場合のテキスト抽出用の一時ファイル作成ディレクトリについて

テキストインデクサーを実行する場合、環境変数ファイルの TMPDIR パラメタに次の一時ファイル作成ディレクトリを必ず作成してください。

- PDF 文書または XML 文書のテキスト抽出用の一時ディレクトリ
- フィルタプラグイン機能でテキストインデクサーとテキスト抽出フィルタプラグインライブラリとのデータの受け渡しをファイルで行う指定の場合の一時ディレクトリ

ディレクトリ名には、存在し、かつテキストインデクサー実行者の書き込み権限があるディレクトリを指定してください。

(10) PDF を示す RenditionType の値に [filterA] または [filterC] のフィルタ拡張子を付けて実行したとき

RenditionType 定義ファイルで、RenditionType の値とフィルタ拡張子の組み合わせが誤った状態でテキストインデクサーを実行すると、不正なテキストをインデクス登録する場合がありますのでご注意ください。RenditionType の指定については、「3.4 RenditionType 定義ファイル」を参照してください。

(11) テキストインデクサーが LimitTime タイムアウトで終了したとき

テキストインデクサーで LimitTime タイムアウトが発生した場合、コマンドは次のように終了します。

- 全文検索用の文書クラスから処理対象文書を取り出す前にタイムアウトが発生した場合、直ちにコマンドを終了します。
- 取り出した文書から、RenditionType 定義情報に基づいてテキストを抽出する処理（テキスト抽出処理）の前にタイムアウトが発生した場合、テキスト抽出処理をキャンセルし、コマンドを終了します。
- テキスト抽出完了後にタイムアウトが発生した場合、テキスト抽出した文書の全文検索インデックスをデータベースに登録し、コマンドを終了します。

タイムアウトが発生する前にメッセージ KMBV12047-I が出力されている文書については正常に登録されます。

(12) テキストインデクサーが対象とする文書クラス、および全文検索インデックスとして指定できるプロパティについて

テキストインデクサーが処理対象にする文書クラス、およびその文書クラスに全文インデックスとして指定できるプロパティについては、「付録 A 処理対象の全文検索インデックス用プロパティ」を参照してください。

(13) 定義ファイル生成コマンドを使用して定義ファイルを生成した場合の注意事項

定義ファイル生成コマンド (EDMCrtTIDF) を使用して定義ファイルを生成すると、半角空白を含むクラス名またはプロパティ名を出力することがあります。テキストインデクサーでは半角空白を含むクラス名またはプロパティ名を使用できませんので、半角空白を含むクラス名またはプロパティ名は半角空白を含まない名前に変更してから使用してください。

(14) テキストインデクサー、DocumentBroker および HiRDB の実行環境

DocumentBroker サーバの DocumentSpace 構成定義ファイル (docspace.ini) に記述した接続情報の定義を使用してデータベースに接続する場合、テキストインデクサーは DocumentBroker サーバと同一のマシン上で動作することが前提となります。

テキストインデクサーの環境設定ファイル (TextIndex.ini) に記述した接続情報の定義を使用してデータベースに接続する場合、テキストインデクサーは DocumentBroker サーバとは別マシンでも動作できます。

(15) Document Filter for Text Search のパス設定について

テキストインデクサーを実行するためには、Document Filter for Text Search のパス設定が必要です。環境変数にパスの設定が行われていることを確認してください。

Windows Server 2003, Windows Server 2003 R2 の場合

「コントロールパネル」にある「システム」の詳細設定タブの「環境変数」ボタンをクリックして表示される「環境変数」ダイアログの「PATH」の設定

2. 実行環境の設定

次のパスを追加してください。

- < Document Filter for Text Search のインストールディレクトリ >
- < Document Filter for Text Search のインストールディレクトリ > %bin

Windows Server 2008 の場合

「コントロールパネル」にある「システム」 - 「システムの詳細設定」の詳細設定タブの「環境変数」ボタンをクリックして表示される「環境変数」ダイアログの「PATH」の設定

次のパスを追加してください。

- < Document Filter for Text Search のインストールディレクトリ >
- < Document Filter for Text Search のインストールディレクトリ > %bin

(16) テキスト抽出時の注意事項

(a) 一時ファイルについて

テキスト抽出処理中にプロセスを強制終了させたり、サポート外の文書ファイルを読み込ませたりすると、一時ディレクトリ内に一時ファイルが残る場合があります。一時ファイルの名前は特に決まっていません。一時ファイルができる場所は、次の順番で決定されます。

1. 環境変数または環境変数ファイル TMPDIR に設定されているディレクトリ
2. インストールディレクトリ下の %tmp

この一時ファイルは、テキスト抽出処理が終われば不要になります。そのままにしておくとディスク容量が圧迫されますので、適宜削除してください。

なお、テキスト抽出できるドキュメントの種類については、各テキスト抽出ライブラリをご確認ください。

(b) テキスト抽出時の現象について

テキスト抽出する文書ファイルによっては、次のような現象が発生することがあります。

文書ファイルのプロパティ情報が空のとき、不要なデータが抽出される場合があります。

抽出したテキストの一部（半角の ¥, 文書を作成するアプリケーションに依存する文字, ユーザ定義文字など）の表示が不正になることがあります。

アプリケーションで変換した文書ファイル（特に Lotus 1-2-3 を Excel 形式で保存した場合）からテキスト抽出すると、抽出したテキストの表示が不正になることがあります。このときは、変換後のアプリケーションで文書ファイルを再保存すると正常にテキスト抽出できるようになります。

図やグラフに含まれるテキストは、抽出できなかつたり、表示が不正になったりすることがあります。

Excel で一部の表が保護されている場合、テキスト抽出に失敗することがあります。

一太郎の文書で、図 2-3 のように、表の一つのセル内にあるデータが複数行で構成されている場合、テキスト抽出時に、「顧客手アンテナ」および「数料工事費」というテキストとして抽出されます。

図 2-3 一つのセル内が複数行で構成された表の例

顧客手 数料	アンテナ 工事費
-----------	-------------

(c) PDF ファイルについて

PDF ファイルによっては次のような現象が発生しますので注意してください。

次のセキュリティが一つでも PDF ファイルに設定されていると、テキストが抽出できません。

- パスワード
- プリント書類の変更
- テキストまたはグラフィックの選択
- ノートの追加または変更

PDF ファイルから抽出されたテキストの順番は、PDF の表示結果と一致しない場合があります。

次の文字の表示が不正になる場合があります。

- 半角のチルダ (~)
- 機種依存文字
- ユーザ定義文字

(17) サブレンディションの文書からのテキスト抽出について

(a) テキストインデクサー実行時の処理

テキストインデクサーの実行 (EDMTextIndex の実行) では、RenditionType 定義ファイルに指定した RenditionType がマスタレンディションの RenditionType と一致しない場合、サブレンディションの RenditionType を検索して、一致した文書に対してテキスト抽出を実行します。

(b) RenditionType 定義ファイルに複数の RenditionType を指定した場合

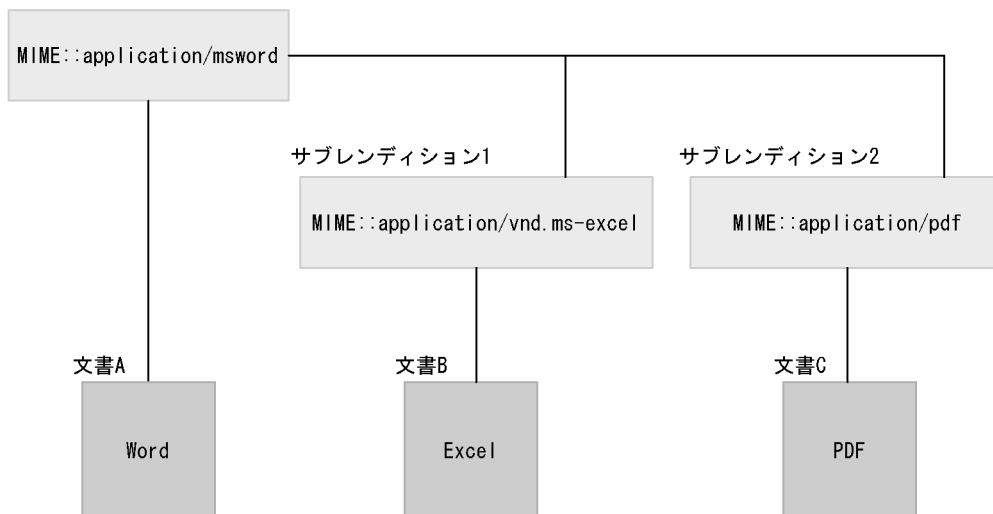
先頭に指定した RenditionType と一致する RenditionType がマスタレンディションおよびサブレンディションに存在しない場合、次に指定した RenditionType と一致する RenditionType をマスタレンディション、サブレンディションの順に検索し、一致した文書に対してテキスト抽出を実行します。

文書オブジェクトの構成例を図 2-4 に示します。また、図 2-4 を例として、RenditionType 定義ファイルの定義とテキストインデクサーが処理対象とする文書との関連について、例 1 ~ 3 で説明します。

2. 実行環境の設定

図 2-4 文書オブジェクトの構成例

マスタレンディション



例 1

```
application/msword:filterA
application/vnd.ms-excel:filterA
application/pdf:filterB
```

この定義例では、一行目に Word の文書が指定されています。マスタレンディションに Word の文書が存在するため、文書 A がテキストインデクサーの処理対象文書になります。

例 2

```
application/pdf:filterB
application/vnd.ms-excel:filterA
application/msword:filterA
```

この定義例では、一行目に PDF の文書が指定されています。マスタレンディション、サブレンディション 1、サブレンディション 2 の順で検索した結果、サブレンディション 2 に PDF の文書が存在するため、文書 C がテキストインデクサーの処理対象文書になります。

例 3

```
application/x-js-taro:filterA
application/vnd.lotus-1-2-3:filterA
application/vnd.ms-excel:filterA
```

この定義例では、一行目に一太郎の文書、二行目に Lotus 1-2-3 の文書が定義されています。それぞれの定義について、マスタレンディション、サブレンディション 1、サブレンディション 2 の順で検索した結果、一太郎の文書および Lotus 1-2-3 の文書は文書オブジェクト中にないため、一行目および二行目の定義に対する処理は実行

されません。

三行目に定義されている Excel の文書が、マスタレンディション、サブレンディション 1 の順で検索した結果、サブレンディション 1 に存在するため、文書 B がテキストインデクサーの処理対象文書になります。

RenditionType 定義ファイルへの RenditionType の指定については、「3.4 RenditionType 定義ファイル」を参照してください。コマンドの詳細については、「4. コマンドリファレンス」を参照してください。

(18) EDMTextIndex コマンド実行時に -o オプションを指定した場合

EDMTextIndex コマンド実行時に -o オプションを指定すると、インデクスの登録状態に関係なく処理対象文書クラス中のすべての文書を処理対象とします。

登録済みのインデクスについては、文書実体ファイルから抽出したインデクスで上書きしますのでご注意ください。

3

テキストインデクサーで使用するファイル

この章では、テキストインデクサーを実行するときに使用する各ファイルの定義方法について説明します。

3.1 使用するファイルの一覧

3.2 ファイル形式と文法

3.3 環境設定ファイル

3.4 RenditionType 定義ファイル

3.5 定義ファイル

3.6 処理対象文書クラス定義ファイル

3.7 環境変数ファイル

3.8 プラグイン環境設定ファイル

3.1 使用するファイルの一覧

この節では、テキストインデクサーで使用するファイルについて説明します。

テキストインデクサーで使用するファイルの一覧を表 3-1 に示します。

表 3-1 テキストインデクサーで使用するファイル

ファイル名称	機能	作成場所
環境設定ファイル (TextIndex.ini)	テキストインデクサー制御用のパラメタを定義するファイルです。このファイルはテキストインデクサーをインストールする際に作成されます。ファイルの設定内容はユーザの環境に合わせて変更する必要があります。なお、テキストインデクサーを運用する際は、全文検索時のデータの整合性をとるために、これらのパラメタを途中で変更しないことを推奨します。	インストールディレクトリ <small>¥etc</small>
RenditionType 定義ファイル (RenditionType.ini)	テキスト抽出ライブラリでテキスト抽出できるアプリケーションプログラムと、レンディションタイプとを対応づけるファイルです。このファイルはテキストインデクサーをインストールする際に作成されます。ファイルの設定内容はカスタマイズできます。	インストールディレクトリ <small>¥etc</small>
環境変数ファイル (NTeonfig.ini)	テキストインデクサーの動作に必要な環境変数を設定するためのファイルです。このファイルは、テキストインデクサーをインストールする際に作成されます。ファイルの設定内容はカスタマイズできます。	インストールディレクトリ <small>¥etc</small>
定義ファイル (ファイル名は任意)	処理対象文書クラスの GUID 値および書誌情報プロパティの GUID 値を取得するために使用するファイルです。作成した定義ファイルのファイル名を環境設定ファイル (TextIndex.ini) の Definition_def エントリに指定して使用します。定義ファイルは [ClassNameDefinition] セクションに処理対象文書クラス名、[PropNameDefinition] セクションに書誌情報プロパティ名を定義します。	任意
処理対象文書クラス定義ファイル (ファイル名は任意)	テキストインデクサー実行時に一つ以上のクラスを処理対象にすると、処理対象にする「処理対象文書クラス名」を指定するファイルです。処理対象文書クラス名とは、定義ファイルの [ClassNameDefinition] セクションに定義したエントリ名のことを指します。	任意
プラグイン環境設定ファイル (PlugInconfig.ini)	次のライブラリを実行するために定義するファイルです。 <ul style="list-style-type: none"> テキストインデクサーが定義するファイル呼び出しインタフェースに従った、テキスト抽出フィルタプラグインライブラリ テキストインデクサーで定義するプロパティ値の取得呼び出しインタフェースに従った、プロパティ値取得用プラグインライブラリ 	インストールディレクトリ <small>¥etc</small>

3.2 ファイル形式と文法

この節では、テキストインデクサーを実行する場合に必要な各種ファイルのファイル形式および文法について説明します。

3.2.1 ファイル形式

テキストインデクサーを実行する場合に必要なファイルの形式は次のとおりです。

- テキストファイル
- 使用できる文字は、OS で扱える文字 (ASCII コード, 日本語は Shift-JIS)

注 文書空間で使用する文字コード種別が UTF-8 の場合, 指定できる文字は印刷可能な ASCII コードだけです。

3.2.2 文法

テキストインデクサーで使用するファイルは、幾つかのセクションとセクションを構成するエントリで構成されます。セクション名およびエントリ名は英字の大文字, 小文字を区別します。したがって、セクション名およびエントリ名は各ファイルの説明に記述してあるとおりの文字列を記述してください。

(1) セクション

セクションを記述する場合は、次の文法に従います。

- セクション名は角括弧 ([]) で囲みます。
- 角括弧とセクション名の間には、半角空白 (0x20) およびタブ (0x09) 以外の文字を入れないでください。

(2) エントリ

セクションを構成するエントリを記述する場合は、次の文法に従います。

- エントリは「エントリ = 値」の形式で、一行で記述します。
- 「=」の前には、半角空白 (0x20) およびタブ (0x09) を挿入できます。
- 行間区切り文字は、<EOF>, <CR>, <LF>, <CR>+<LF> です。
- 一行は、半角空白および行間区切り文字を含めて 8,192 バイト以内で記述します。

(3) パスの指定方法

パスとは、ディレクトリ名の文字列にファイル名の文字列を追加した文字列を示します。次の規則に従って指定してください。

指定規則

- 指定できる文字列は、Windows で作成できるディレクトリ名やファイル名の文字

3. テキストインデクサーで使用するファイル

列です。ただし、文書空間で使用する文字コード種別が UTF-8 の場合、指定できる文字は印刷可能な ASCII コードだけです。

- 指定できるパスの長さは 255 バイトです。大文字と小文字は区別されません。
- パスは、ディレクトリ名が「<ドライブ指定文字>:」で始まる場合は絶対パス、それ以外の文字で始まる場合は相対パスとみなされます。
- パスに特殊文字を含める場合は、必ず半角の引用符 (") でパスを囲んでください。
- 特殊文字の定義は Windows での定義に従います。例えば、空白を含むディレクトリ名を作成する場合は、ディレクトリ作成コマンドに続いてディレクトリ名を半角の引用符 (") で囲みます。このように Windows で扱える文字のうちで、「"」で囲んで扱う必要のある文字を特殊文字にします。

指定例

パスの指定例を次に示します。

C:¥TMP¥InputDataFileName.txt	←絶対パスの指定
.¥InputDataFileName.txt	←相対パスの指定
InputDataFileName.txt	←相対パスの指定
"Doc Broker¥DATA¥InputDataFileName.txt"	←パスに特殊文字を含む場合

(4) その他

- 先頭の 1 カラム目が「#」の行はコメント行になります。

3.3 環境設定ファイル

この節では、環境設定ファイルについて説明します。

3.3.1 記述形式

環境設定ファイル (TextIndex.ini) の記述形式を図 3-1 に示します。

図 3-1 環境設定ファイルの記述形式

```
#△コメント行△<改行>
パラメタ名△=△各パラメタの設定値△<改行>

(凡例)
△: 0個以上の (半角空白またはタブ)
#: コメント行の開始記号
<改行>: 行間区切り文字
```

3.3.2 記述規則

環境設定ファイルの記述規則を次に示します。

規則

1. 各定義は 1 行で記述します。
2. 1 文字目から記述します。
3. 文字列の後ろに半角空白またはタブを含めることができます。
4. 「=」の前後には、半角空白またはタブを挿入できます。
5. 使用できる改行コードは (0x0d0a) です。
6. 英字の大文字と小文字は区別されます。
7. シャープ (#) で始まる行はコメント行になります。
8. 同じエントリが複数存在する場合は最上位の指定が有効になります (ワーニングメッセージ KMBV12023-W が出力されます)。

注意事項

- 環境設定ファイル中に不正な文字列がある場合は、エラーメッセージ KMBV12012-E を出力して処理を終了します。

3.3.3 環境定義パラメタ

環境設定ファイルに記述する環境定義パラメタを次に示します。必要なパラメタを TextIndex.ini ファイルに記述してください。

(1) Definition_def

形式

3. テキストインデクサーで使用するファイル

Definition_def = 定義ファイル名

指定方法

1. 定義ファイルのファイル名を指定します。
定義ファイルは、処理対象文書クラスおよび書誌情報プロパティを「エントリ名=GUID 値」の形式で定義したファイルです。
2. このパラメタは省略できません。
3. ファイル名に空白が含まれる場合は " で囲む必要があります。

(2) RenditionType_def

形式

RenditionType_def = 定義ファイル名

指定方法

1. 対応する RenditionType 定義ファイルのファイル名を指定します。
2. このパラメタは省略できません。
3. ファイル名に空白が含まれる場合は " で囲む必要があります。

(3) MaxOfLength

形式

MaxOfLength = テキスト抽出長の上限值

指定方法

1. 文書ごとのテキスト抽出長の上限值 (単位: バイト) (0 ~ 5242780) を指定します。
2. 上限値に 0 を指定するとテキストインデクサーは何も処理しません。
3. 指定内容が正しくない場合は、エラーメッセージ KMBV12015-E が出力されます。
4. テキスト抽出結果が指定した値を超えた場合、メッセージ KMBV12009-I が出力され、MaxOfLength で指定した長さまでのテキストが抽出されます。ただし、XML 文書の場合は、KMBV12009-I の出力後に KMBV12037-W が出力され、MaxOfLength で指定した分のテキスト抽出結果は登録されません。テキスト抽出結果については、「2.6 実行環境および運用に関する注意事項」を参照してください。
5. このパラメタは省略できます。省略すると、5242780 が仮定されます。

注意事項

- テキストインデクサーは MaxOfLength を超えた時点で文書の処理を終了しない場合があります。これは文書および文書内容によって異なります。したがって、MaxOfLength では、実行時間の制限はできません。

(4) LimitTime

形式

LimitTime = 実行時間上限値

指定方法

1. テキストインデクサーの実行時間の上限値（単位：秒）（1 ~ 2147483647）を指定します。
2. このパラメタは省略できます。省略すると、実行時間を監視しません。

(5) ModeOfProperty

形式

ModeOfProperty = [ON | OFF]

指定方法

1. コンテンツ取得エラー発生時の書誌情報取得オプション [ON | OFF] を指定します。
2. このパラメタは省略できます。省略すると、「OFF」が仮定されます。

(6) PropertyOfIndex

形式

PropertyOfIndex = 書誌情報プロパティ [, 書誌情報プロパティ...]

指定方法

1. ModeOfProperty=ON を指定したときに取得する書誌情報プロパティを指定します。
2. 書誌情報プロパティにはユーザが処理対象文書クラスに設定した文書オブジェクトのユーザプロパティの名称を、コンマ(,)区切りで指定します。
3. 指定できる文字列の長さは、128 バイト以下です。
4. 指定できる文書オブジェクトのユーザプロパティのデータ型は、INTEGER 型、CHAR 型、MVARCHAR 型の 3 種類です。
5. このパラメタは省略できます。省略すると、書誌情報を取得しません。

(7) ErrorLog

形式

ErrorLog = エラーログファイルの出力先パス

指定方法

1. エラーログファイルの出力先パスを指定します。
2. このパラメタは省略できます。省略すると、エラーログファイルは作成されません。
3. 指定内容が正しくない場合は、エラーメッセージ KMBV12004-E が出力されます。
4. エラーログファイルの出力先パスに空白が含まれる場合は " で囲む必要があります。

注意事項

- エラーログファイルのセキュリティで設定するアクセス権として、「読み取り」「書き込み」「削除」が必要です。

3. テキストインデクサーで使用するファイル

- エラーログファイルを格納するディレクトリのセキュリティで設定するアクセス権として、「読み取り」「書き込み」「実行」が必要です。
- エラーログファイルを格納するディレクトリ下のファイルのセキュリティで設定するアクセス権として、「読み取り」「書き込み」「削除」が必要です。

(8) XmlBroker

形式

XmlBroker = HAX

指定方法

1. XML 文書から全文検索インデクスを登録する場合、使用するソフトウェアを指定します。
HAX : HiRDB Adapter for XML を使用します。
2. このパラメタは省略できません。
3. 不正な値（「HAX」以外の値）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(9) DocSpace_def

形式

DocSpace_def = [ON | OFF]

指定方法

1. データベースへの接続方法を指定します。
ON : テキストインデクサーの接続情報の定義を使用して、データベースに接続します。
OFF : DocumentBroker サーバの接続情報の定義を使用して、データベースに接続します。
2. このパラメタは省略できます。省略した場合、「OFF」が仮定されます。
3. 不正な値（「ON」「OFF」以外の値）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

注意事項

- DocSpace_def = ON を指定した場合、インデクサーは DocumentBroker サーバのデータベース運用コマンドとデータベース移行コマンドとの間の排他制御を実行しません。このため、両コマンドを同時に実行しないように注意してください。

(10) SystemObjectID

形式

SystemObjectID = DMA SystemオブジェクトのOID (シリアルID)

指定方法

1. DocumentBroker サーバの smgrreg.ini に定義されている、ServiceObjectID エントリのエントリ値「07a17522-a626-11d0-b11f-0020af27a837」を指定します。

2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な値（「8 けた -4 けた -4 けた -4 けた -12 けた」以外の形式）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(11) DocSpaceId

形式

DocSpaceId = 文書空間識別子

指定方法

1. 文書空間識別子を、「8 けた -4 けた -4 けた -4 けた -12 けた」の形式で指定します。
2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な形式（「8 けた -4 けた -4 けた -4 けた -12 けた」以外の形式）の値を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(12) DbType

形式

DbType = HIRDB

指定方法

1. 接続するデータベースの種別として「HIRDB」を指定します。
2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な値（「HIRDB」以外の値）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(13) PdHost

形式

PdHost = HiRDBサーバのホスト名

指定方法

1. HiRDB サーバのホスト名を 1 ~ 256 バイトで指定します。
2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な値（1 ~ 256 バイトの範囲を超えた値）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(14) PdNamePort

形式

PdNamePort = HiRDBサーバのポート番号

3. テキストインデクサーで使用するファイル

指定方法

1. HiRDB サーバのポート番号を 1 ~ 2147483647 の範囲で指定します。
2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な値 (1 ~ 2147483647 の範囲を超えた値) を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(15) PdUser

形式

PdUser = { "ユーザ名" | ユーザ名 } [/ { "パスワード" | パスワード }]

指定方法

1. HiRDB にアクセスするためのユーザ名およびパスワードを、それぞれ 1 ~ 30 バイトの範囲で指定します。このとき、ユーザ名には、HiRDB のユーザ権限で CONNECT 権限およびスキーマ定義権限を付与したユーザを指定してください。
2. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
3. 不正な形式の値を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(16) PdTSPluginOwner

形式

PdTSPluginOwner = { "ユーザの認可識別子" | ユーザの認可識別子 }

指定方法

1. HiRDB Text Search Plug-in を登録したユーザの認可識別子を、1 ~ 30 バイトで指定します。
2. このパラメータは、HiRDB Text Search Plug-in の登録時に、pdplrgst コマンドの -u オプション (プラグイン登録スキーマ指定) を指定したときだけ指定してください。
3. DocSpace_def の指定が ON の場合は、この指定を省略できません。DocSpace_def の指定が OFF の場合は、この指定は無視されます。
4. 不正な形式の値を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

(17) ModeOfIndexing

形式

ModeOfIndexing = [NOLOCK | CONSISTENT]

指定方法

1. 全文検索インデクス登録処理時のテキストインデクサーの動作を指定します。
NOLOCK : 文書ファイルと全文検索インデクスの内容が一致しているかを確認し

ないで、全文検索インデクスを登録します。

CONSISTENT：文書ファイルと全文検索インデクスの内容が一致しているかを確認し、一致する場合だけ全文検索インデクスを登録します。

2. このパラメタは省略できます。省略した場合、「NOLOCK」が仮定されます。
3. 不正な値（「NOLOCK」「CONSISTENT」以外の値）を指定した場合は、エラーメッセージ KMBV12012-E が出力されます。

注意事項

- ModeOfIndexing = NOLOCK を指定した場合、文書ファイルと全文検索インデクスの内容が一致しているかを確認しないため、「CONSISTENT」を指定した場合よりも高速に全文検索インデクスの登録処理ができます。
- ModeOfIndexing = CONSISTENT を指定した場合、文書ファイルと全文検索インデクスの内容が一致しないときは、全文検索インデクスの登録処理が中止されてエラーメッセージ KMBV12079-W が出力されます。

3.3.4 抽出処理が失敗した文書の取り扱い

環境定義パラメタのオプション設定が ModeOfProperty = ON であれば、コンテンツからテキスト抽出ができなかった場合でも、環境定義パラメタ (PropertyOfIndex) に指定された書誌情報を取得して、これをインデクスファイルとして登録し、全文検索時に利用できます。ModeOfProperty = OFF の場合は、インデクスファイルが登録されません。

書誌情報とは文書に対する付加情報であり、一般的に文書のタイトル、作成者、作成日時などの情報が書誌情報として使用されます。

インデクスファイルに出力される書誌情報の形式は、PropertyOfIndex に指定された文書プロパティから取得したデータを、半角スペース区切りで連結したものです。書誌情報として、環境定義パラメタ (PropertyOfIndex) にタイトル、作者、コメントが指定されていた場合の書誌情報の作成例を図 3-2 に示します。

図 3-2 書誌情報の作成例

テキストインデクサー機能仕様書△日立 太郎△1999年7月17日第1版作成
(凡例)
△：0個以上の（半角空白またはタブ）

文書プロパティから取得したデータ中に ¥0 (0x00) が存在した場合、書誌情報は連結後の先頭から ¥0 までの文字列になります。ただし、書誌情報の先頭に ¥0 が存在した場合は HiRDB のメッセージが出力され、書誌情報は登録されません。

3.3.5 環境設定ファイルの記述例

環境設定ファイル (TextIndex.ini) の記述例を示します。

3. テキストインデクサーで使用するファイル

```
Definition_def = C:\etc\EDM_lod_define.txt
RenditionType_def = C:\etc\RenditionType.ini
MaxOfLength = 5242780
LimitTime = 3600
ModeOfProperty = OFF
PropertyOfIndex = D-TITLE, D-COMMENT, D-AUTHOR
XmlBroker = HAX
DocSpace_def = ON
SystemObjectId = 07a17522-a626-11d0-b11f-0020af27a837
DocSpaceId = 673d2be0-d1fd-11d0-ab59-08002be29e1d
DbType = HIRDB
PdHost = hostname
PdNamePort = 20550
PdUser = "user"/"pass"
ModeOfIndexing = NOLOCK
```

3.4 RenditionType 定義ファイル

この節では、RenditionType 定義ファイルについて説明します。

3.4.1 提供するフィルタ拡張子と使用するテキスト抽出ライブラリ

RenditionType 定義ファイルは、RenditionType に「:フィルタ拡張子」を付けて定義します。表 3-2 に提供するフィルタ拡張子と使用するテキスト抽出ライブラリを示します。テキスト抽出できるドキュメントの種類については、各テキスト抽出ライブラリをご確認ください。

表 3-2 提供するフィルタ拡張子と使用するテキスト抽出ライブラリ

フィルタ拡張子	テキスト抽出ライブラリ
filterA	Document Filter for Text Search
filterB	Document Filter for Text Search
filterC	- 1
filterD	環境設定ファイル (TextIndex.ini) の XmlBroker エントリに指定したテキスト抽出ライブラリ
filterE_XXXX ²	テキスト抽出フィルタプラグインライブラリ

注

フィルタ拡張子とドキュメントの種類の組み合わせが誤った状態でテキストインデクサーを実行すると、不正なテキストをインデクス登録する場合がありますのでご注意ください。

注 1

テキスト抽出ライブラリを使用しないで、テキストファイルを全文検索インデクスに登録します。

注 2

フィルタプラグイン用のフィルタ拡張子です。XXXX には、0000 ~ 9999 の数値を指定します。このフィルタ拡張子を指定する場合は、必ずプラグイン環境設定ファイルに [filterE_XXXX] セクションを定義してください。[filterE_XXXX] セクションを定義していない場合は、エラーメッセージ KMBV12206-E が出力されます。

3.4.2 記述形式

RenditionType 定義ファイル (RenditionType.ini) の記述形式を図 3-3 に示します。

3. テキストインデクサーで使用するファイル

図 3-3 RenditionType 定義ファイルの記述形式

```
#△コメント行△<改行>
RenditionType△[:△パラメタ][△(コメント)]△:△フィルタ拡張子△<改行>

(凡例)
# : コメント行の開始記号
△ : 0個以上の (半角空白またはタブ)
<改行> : 行間区切り文字
```

3.4.3 記述規則

RenditionType 定義ファイルの記述規則を次に示します。

規則

1. 各定義は 1 行で記述します。
2. 1 文字目から記述します。
3. 文字列の後ろに半角空白またはタブを含めることができます。
4. 各項目の間には、半角空白またはタブを挿入できます。
5. 使用できる改行コードは (0x0d0a) です。
6. 英字の大文字と小文字は区別されます。
7. シャープ (#) で始まる行はコメント行になります。
8. RenditionType 名の長さは 255 バイト以内です。
9. RenditionType 名の先頭に「MIME::」は付けしないでください。処理対象クラスの RenditionType プロパティから「MIME::」を取り除いたものを使用します。
10. 制御コード値 (0x00 ~ 0x1F) は RenditionType 名に使用できません。
11. パラメタやコメントを指定しなかった場合は、パラメタやコメントが指定されていない RenditionType 名が対象になります。
12. RenditionType 名の先頭以外の任意の位置でワイルドカードが使用できます。ワイルドカード文字以降に半角空白またはタブ以外の文字を記述した場合は、ワイルドカード文字は通常の文字になります。使用できるワイルドカードを表 3-3 に示します。

表 3-3 使用できるワイルドカード

項番	文字	意味
1	* (半角アスタリスク)	0 文字以上の任意の文字列

ワイルドカードの記述例を次に示します。

```
#RenditionType以降任意の文字列
RenditionType*:filterA
#パラメタ"parame"以降任意の文字列
```

```

RenditionType ;parame*:filterA
#コメント" ("以降任意の文字列
RenditionType (*:filterA
#ワイルドカード文字以降に半角空白またはタブ以外の文字を記述した場合
RenditionType*HOGE:filterA
# 「RenditionType*HOGE」という値のRenditionTypeが対象になります。

```

注意事項

- RenditionType 定義ファイルに同じ RenditionType を指定する RenditionType 名が複数存在する場合はワーニングメッセージ KMBV12024-W を出力します。
- RenditionType 名の記述が正しくない場合は、エラーメッセージ KMBV12028-E を出力して処理を終了します。
- RenditionType 名を追加する場合は、処理対象文書クラスの RenditionType プロパティに設定されている RenditionType 名を参照してこれを使用する必要があります。
- テキスト抽出ライブラリがサポートしていないレンディションタイプの文書が、処理対象クラスに存在する場合、その文書は処理されないで、ワーニングメッセージ KMBV12045-W を出力します。

3.4.4 RenditionType 定義ファイルの記述例

RenditionType 定義ファイル (RenditionType.ini) の記述例を次に示します。

```

# アプリケーションに対応するレンディションタイプ定義
application/rtf:filterA
application/msword*:filterA
application/vnd.ms-excel;charset=ascii (EXCEL):filterA
application/ms-powerpoint:filterA
application/x-js-taro:filterA
application/vnd.lotus-1-2-3:filterA
application/pdf:filterB
text/plain:filterC
text/xml:filterD
application/tif:filterE_0001
# 定義ファイル END

```

3.5 定義ファイル

この節では、定義ファイルについて説明します。

3.5.1 記述形式

定義ファイルの記述形式を図 3-4 に示します。

図 3-4 定義ファイルの記述形式

```
[ClassNameDefinition]
処理対象文書クラス名△=△GUID値△<改行>
処理対象文書クラス名△=△GUID値△<改行>
:
[PropNameDefinition]
書誌情報プロパティ名△=△GUID値△<改行>
書誌情報プロパティ名△=△GUID値△<改行>
:
(凡例)
△: 0個以上の (半角空白またはタブ)
<改行>: 行間区切り文字
```

3.5.2 記述規則

定義ファイルの記述規則を次に示します。

規則

1. セクション名や各エントリの定義は 1 行で記述します。
2. 1 文字目から記述します。
3. 文字列の後ろに半角空白またはタブを含めることができます。
4. 使用できる改行コードは (0x0d0a) です。
5. 英字の大文字と小文字は区別されます。
6. シャープ (#) で始まる行はコメント行になります。
7. 処理対象文書クラス名と書誌情報プロパティ名の長さは 128 バイト以内です。
8. 処理対象文書クラス名と書誌情報プロパティ名には先頭が「**」で始まる名称は指定できません。
9. 処理対象文書クラス名と書誌情報プロパティ名には "@", "=", ".", "," と半角空白は使用できません。
10. GUID 値は "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx" の形式で指定します。x は 0 ~ 9, a ~ f の文字です。

(1) [ClassNameDefinition] セクション

[ClassNameDefinition] セクションは、処理対象文書クラスの GUID 値を取得するために使用します。

図 3-5 の形式で処理対象文書クラスを定義します。

図 3-5 定義ファイルの処理対象文書クラスの定義方法

```
処理対象文書クラス=GUID値<改行>
```

(凡例)

<改行> : 行間区切り文字

(2) [PropNameDefinition] セクション

[PropNameDefinition] セクションは書誌情報プロパティの GUID 値を取得するために使用します。

図 3-6 の形式で書誌情報プロパティを定義します。

図 3-6 定義ファイルの書誌情報プロパティの定義方法

```
書誌情報プロパティ=GUID値<改行>
```

(凡例)

<改行> : 行間区切り文字

3.5.3 定義ファイルの記述例

定義ファイルの記述例を次に示します。

```
[ClassNameDefinition]
```

```
Class_ContentSearchDV1=76fa6abe-0406-11d2-b29a-0060b0ea4840
```

```
Class_ContentSearchDV2=76fa595c-0406-11d2-b29a-0060b0ea4840
```

```
[PropNameDefinition]
```

```
D-TITLE=76fa162c-0406-11d2-b29a-0060b0ea4840
```

```
D-COMMENT=76fa3440-0406-11d2-b29a-0060b0ea4840
```

```
D-AUTHOR=76fa5222-0406-11d2-b29a-0060b0ea4840
```

3.6 処理対象文書クラス定義ファイル

この節では、処理対象文書クラス定義ファイルについて説明します。

3.6.1 記述形式

処理対象文書クラス定義ファイルの記述形式を図 3-7 に示します。

図 3-7 処理対象文書クラス定義ファイルの記述形式

```
処理対象文書クラス名△<改行>
処理対象文書クラス名△<改行>
      :
#コメント△<改行>

(凡例)
  △ : 0個以上の (半角空白またはタブ)
  # : コメント行の開始記号
<改行> : 行間区切り文字
```

3.6.2 記述規則

処理対象文書クラス定義ファイルの記述規則を次に示します。

規則

1. 各定義は 1 行で記述します。
2. 1 文字目から記述します。
3. 文字列の後ろに半角空白またはタブを含めることができます。
4. 使用できる改行コードは (0x0d0a) です。
5. 英字の大文字と小文字は区別されます。
6. シャープ (#) で始まる行はコメント行になります。
7. クラス名の長さは 128 バイト以内です。
8. 文字列の途中に半角空白を含めることはできません。
9. 処理対象文書クラスには、DocumentBroker で登録されている全文検索対象クラスを指定します。

3.6.3 処理対象文書クラス定義ファイルの記述例

処理対象文書クラス定義ファイルの記述例を次に示します。

```
# 処理対象文書クラスの定義
Class_ContentSearchDV1
Class_ContentSearchDV2
      :
      :
```

定義ファイルEND

3.7 環境変数ファイル

この節では、環境変数ファイルについて説明します。

環境変数ファイルは、テキストインデクサーが動作するために必要な情報を設定するファイルです。

3.7.1 記述形式

環境変数ファイルの記述形式を図 3-8 に示します。

図 3-8 環境変数ファイル (NTconfig.ini) の記述形式

```
[ConfigIndex]
_HIEDMS_TRACE_DIR△=△トレースログファイルの出力先ディレクトリのパス名△<改行>
_HIEDMS_TRACE_NUM△=△トレースログファイルのファイル数△<改行>
_HIEDMS_TRACE_SIZE△=△トレースログファイルのファイルサイズ△<改行>
_HIEDMS_TRACE_LEVEL△=△トレースレベル△<改行>
TMPDIR△=△一時ディレクトリのパス名*△<改行>
```

(凡例)

△ : 0個以上の (半角空白またはタブ)

<改行> : 行間区切り文字

注※ 次のどちらかを指定する。

- ・PDF文書またはXML文書のテキスト抽出用一時ディレクトリのパス名
- ・フィルタプラグイン機能でテキストインデクサーとテキスト抽出フィルタプラグインライブラリとのデータの受け渡しをファイルで行う指定の場合の一時ディレクトリのパス名

3.7.2 記述規則

環境変数ファイルの記述規則を次に示します。

規則

1. 各定義は 1 行で記述します。
2. 1 文字目から記述します。
3. 文字列の後ろに半角空白またはタブを含めることができます。
4. 使用できる改行コードは (0x0d0a) です。
5. 英字の大文字と小文字は区別されます。
6. シャープ (#) で始まる行はコメント行になります。
7. 文字列の途中に半角空白を含めることはできません。
8. 同じエントリが複数存在する場合は最上位の指定が有効になります (ワーニングメッセージ KMBV12023-W が出力されます)。
9. [ConfigIndex] セクションのあとに記述されたエントリが有効になります。それ以外の場所に記述された文字列はコメントとして扱われます。
10. 指定できるエントリ以外の文字列はコメントとして扱われます。

11. NTconfig.ini にアクセス権限がない場合、NTconfig.ini ファイルがない場合、またはパス名が 256 バイト以上の場合は、エラーメッセージ KMBV20002-E を出力して処理を終了します。

注意事項

インストール直後の NTconfig.ini の初期状態では、パラメタの設定内容はコメント行になっています。

3.7.3 環境変数ファイルに設定するパラメタ

環境変数ファイルに設定するパラメタを次に示します。必要なパラメタを NTconfig.ini ファイルに記述してください。

(1) _HIEDMS_TRACE_DIR

形式

_HIEDMS_TRACE_DIR = トレースログファイルの出力先ディレクトリのパス名

指定方法

1. トレースログファイルの出力先ディレクトリのパス名を設定します。設定したディレクトリの下に ¥index がトレースログファイルの出力先となります。
2. このパラメタを省略した場合、インストールディレクトリ下の ¥spool¥index が出力先ディレクトリになります。
3. _HIEDMS_TRACE_DIR に設定したディレクトリが存在しない場合、または 256 バイト以上の場合には、ワーニングメッセージ KMBV20000-W を出力し、インストールディレクトリ下の ¥spool¥index を使用します。

(2) _HIEDMS_TRACE_NUM

形式

_HIEDMS_TRACE_NUM = トレースログファイルのファイル数

指定方法

1. トレースの出力情報がトレースファイルのサイズの上限を超えた場合、トレースの出力先を別のファイルに切り替えられます。この場合の切り替えるファイル数を設定します。
2. 設定できるファイルの数は 2 ~ 16 の値です。デフォルトは 2 です。
3. 設定したファイル数までファイルが切り替わると、そのあとは最初に使用したファイルから順に上書きされます。

(3) _HIEDMS_TRACE_SIZE

形式

_HIEDMS_TRACE_SIZE = トレースログファイルのファイルサイズ

指定方法

3. テキストインデクサーで使用するファイル

トレースを出力するファイルのサイズを 4096 ~ 2147483647 (バイト) で設定します。デフォルトは 1048576 (1 メガバイト) です。

(4) _HIEDMS_TRACE_LEVEL

形式

`_HIEDMS_TRACE_LEVEL =` トレースレベル

指定方法

トレースの出力レベルを設定します。出力レベルには、0 または 10 が設定できます。デフォルトは 10 です。

それぞれを設定した場合に出力される情報を次に示します。

0 を設定した場合

- エラー情報
- サーバの開始と終了
- データベースのエラー情報

10 を設定した場合

- トレースレベルが 0 の場合に出力される情報
- ユーザーインターフェースの情報
- 他のプログラムとのインターフェースの情報
- データベースのコール、リターン情報、発行した SQL 情報

(5) TMPDIR

形式

`TMPDIR =` { PDF文書のテキスト抽出用一時ディレクトリのパス名 |
XML文書のテキスト抽出用一時ディレクトリのパス名 |
フィルタプラグイン機能でテキストインデクサーとテキスト抽出
フィルタプラグインライブラリとのデータの受け渡しをファイルで
行う指定の場合の一時ディレクトリのパス名 }

指定方法

1. 次のパス名を指定します。

PDF 文書または XML 文書のテキスト抽出用一時ディレクトリのパス名

フィルタプラグイン機能でテキストインデクサーとテキスト抽出フィルタプラグインライブラリとのデータの受け渡しをファイルで行う指定の場合の一時ディレクトリのパス名

2. このパラメタを省略した場合、インストールディレクトリ下の `¥tmp` が一時ディレクトリになります。
3. `TMPDIR` に指定したディレクトリが存在しない場合、または 256 バイト以上の場合には、ワーニングメッセージ `KMBV20000-W` を出力し、インストールディレクトリ下の `¥tmp` を使用します。

注意事項

- 一時ディレクトリのセキュリティで設定するアクセス権として、「読み取り」「書

- き込み」「実行」が必要です。
- 一時ディレクトリ下のファイルのセキュリティで設定するアクセス権として、「読み取り」「書き込み」「削除」が必要です。

3.8 プラグイン環境設定ファイル

この節では、プラグイン環境設定ファイルの定義方法について説明します。

3.8.1 記述形式

プラグイン環境設定ファイルの記述形式を次に示します。

図 3-9 プラグイン環境設定ファイル (PlugInconfig.ini) の記述形式

```
[filterE_XXXX] △<改行>
Library△=△XXX△<改行>

#CCC△<改行>

[function] △<改行>
RenditionType△=△XXX△<改行>
Library△=△XXX△<改行>
Function△=△XXX△<改行>
Argv△=△XXX△<改行>
```

(凡例)

△ : 0個以上の (半角空白またはタブ)

: コメント行の開始記号

<改行> : 行間区切り文字

3.8.2 記述規則

プラグイン環境設定ファイルの記述規則を次に示します。

規則

1. セクション名やエントリは 1 行で記述します。
2. 文字列の後ろに半角空白またはタブを含めることができます。
3. 使用できる改行コードは (0x0d0a) です。
4. 英字の大文字と小文字は区別されます。
5. シャープ (#) で始まる行はコメント行になります。
6. 一つのセクションに複数のエントリを指定できるエントリ以外は、同じエントリが複数存在する場合には、最上位の指定が有効になります (ワーニングメッセージ KMBV12207-W が出力されます)。
7. 同じ RenditionType エントリの値が複数のセクションで存在する場合は、最上位の指定が有効になります (ワーニングメッセージ KMBV12207-W が出力されません)。
8. 同じ [filterE_XXXX] セクションが複数存在する場合は、最上位の指定が有効になります (ワーニングメッセージ KMBV12208-W が出力されます)。例えば、filterE_0001 を二つ指定した場合は、最初に指定した filterE_0001 が有効になります。filterE_0001 と filterE_0002 を指定した場合は、どちらも有効です。

9. [filterE_XXXX] セクションの XXXX は、0000 ~ 9999 の範囲で指定します。
この範囲を超えた値を指定した場合はエラーとなります（エラーメッセージ KMBV12203-E が出力されます）
10. フィルタプラグイン機能を使用しないときでも、次の場合にはエラーとなります（エラーメッセージ KMBV12202-E が出力されます）
- プラグイン環境設定ファイルを削除した場合
 - プラグイン環境設定ファイルに読み取り権限がない場合
 - 0 バイトのファイルを作成した場合

(1) [filterE_XXXX] セクション

[filterE_XXXX] セクション名には、RenditionType エントリの値に一致する文書のテキスト抽出後に定義ファイルで指定したフィルタ拡張子を指定します。このセクションには、次のエントリを定義します。

(a) Library

形式

Library = テキスト抽出フィルタプラグインライブラリ

指定方法

1. テキストインデクサーが定義するフィルタ呼び出しインタフェースに従ったテキスト抽出フィルタプラグインライブラリを、255 バイト以内のフルパス名で指定します。
2. 半角空白を含むパス名を指定する場合は、パス名を " で囲んでください。
3. この指定は省略できません。
4. テキスト抽出フィルタプラグインライブラリを 256 バイト以上で指定した場合は、エラーメッセージ KMBV12205-E が出力されます。
5. このエントリに指定したテキスト抽出フィルタプラグインライブラリおよびフィルタプラグイン関数 (dbrIdxTextFilter) のローディングに失敗した場合は、ワーニングメッセージ KMBV12252-W またはエラーメッセージ KMBV12308-E が出力されます。

(b) DataMode

形式

DataMode = [MEMORY | FILE]

指定方法

1. テキストインデクサーとテキスト抽出フィルタプラグインライブラリとのデータの受け渡し方法を指定します。
MEMORY : メモリ上でデータを受け渡す
FILE : ファイルでデータを受け渡す
2. この指定は省略できます。省略した場合は、「MEMORY」が仮定されます。
3. 不正な値（「MEMORY」および「FILE」以外の値）を指定した場合は、エラー

3. テキストインデクサーで使用するファイル

メッセージ KMBV12205-E を出力します。

(2) [function] セクション

[function] セクションには、RenditionType エントリに指定した文書のテキスト抽出後にプロパティ値取得用プラグインライブラリを動作させる場合に指定します。このセクションは、複数指定できます。

指定するエントリを次に示します。

(a) RenditionType

形式

RenditionType = RenditionType名

指定方法

1. 処理対象クラスの RenditionType プロパティから「MIME::」を取り除いたものを、255 バイト以内で指定します。
2. この指定は省略できません。
3. 不正な値（256 バイト以上の名称）を指定した場合は、エラーメッセージ KMBV12205-E を出力します。

(b) Library

形式

Library = プロパティ値取得用プラグインライブラリ

指定方法

1. テキストインデクサーが定義するプロパティ値の取得呼び出しインタフェースに従った、プロパティ値取得用プラグインライブラリを、255 バイト以内のフルパス名で指定します。
2. パス名に半角空白文字を含む場合は、パス名をダブルクォーテーションで囲んで指定してください。このとき、パス名はダブルクォーテーションを含めなくて 255 バイト以内としてください。
3. この指定は省略できません。
4. 不正な値（256 バイト以上のパス名）を指定した場合は、エラーメッセージ KMBV12005-E を出力します。
5. 指定したプロパティ値取得用プラグインライブラリのローディングに失敗した場合は、ワーニングメッセージ KMBV12254-W またはエラーメッセージ KMBV12309-E が出力されます。

(c) Function

形式

Function = 関数名

指定方法

1. テキストインデクサーが定義するプロパティ値の取得呼び出しインタフェースに従った関数の関数名を、128 バイト以内で指定します。
2. この指定は省略できません。
3. 不正な値（129 バイト以上の関数名）を指定した場合は、エラーメッセージ KMBV12205-E を出力します。

(d) Argv

形式

Argv = 文書オブジェクトのユーザプロパティのプロパティ名

指定方法

1. 文書オブジェクトのユーザプロパティのプロパティ名を 128 バイト以内で指定します。
2. 指定できるプロパティのデータ型は、String、Integer32、Boolean の 3 種類です。
3. 複数のプロパティを取得する場合は、取得したい数だけ Argv エントリを指定してください。
4. この指定は省略できません。
5. 129 バイト以上のプロパティ名を指定した場合はエラーメッセージ KMBV12205-E を、存在しないプロパティ名を指定した場合はエラーメッセージ KMBV12009-E を出力します。

3.8.3 プラグイン環境設定ファイルの記述例

プラグイン環境設定ファイルの記述例を次に示します。

```
# Filter Plug-In Definition
[filterE_0001]
Library=/opt/HiEDMS/IndexLoader/sample/libfilter_1.sl
DataMode=MEMORY
[filterE_0002]
Library=/opt/HiEDMS/IndexLoader/sample/libfilter_2.sl
DataMode=FILE

# Get Property Definition
[function]
RenditionType=application/msword
Library=/opt/HiEDMS/IndexLoader/sample/libfilter_1.sl
Function=usrGertProp_1
Argv=usrProp_1
[function]
RenditionType=application/tif
Library=/opt/HiEDMS/IndexLoader/sample/libfilter_2.sl
Function=usrGertProp_2
Argv=usrProp_2
Argv=usrProp_3
```


4

コマンドリファレンス

この章では、テキストインデクサーで使用するコマンドの形式と文法について説明します。

実行コマンド一覧

コマンドの形式

EDMTextIndex (テキストインデクサーの実行)

EDMCrtTIDF (テキストインデクサー定義ファイル生成)

実行コマンド一覧

テキストインデクサーが提供するコマンドの一覧を表 4-1 に示します。

表 4-1 テキストインデクサーが提供するコマンドの一覧

コマンド名	説明
EDMTextIndex	テキストインデクサーを実行します。
EDMCrtTIDF	定義ファイルを生成します。

コマンドの形式

テキストインデクサーで使用するコマンドの入力形式、使用方法および注意事項について説明します。

入力形式

コマンドの入力形式を次に示します。

コマンド名称〔オプション...〕

コマンド名称

コマンド名称は、実行するコマンドのファイル名です。

オプション

オプションの入力形式の規則を次に示します。なお、説明文で使用する「\$」はコマンドプロンプト、「cmd」はコマンド名を表します。

オプションの形式

オプションはハイフン (-) で始まる文字列で、次に示すように、引数を指定しないかまたは 1 個の引数を指定します。

形式 1: - オプションフラグ

形式 2: - オプションフラグ <空白またはタブ> フラグ引数

(凡例)

オプションフラグ: 1 文字の英数字で、英大文字・小文字は区別されます。

フラグ引数: オプションフラグに対する引数です。

オプションの指定規則

- フラグ引数をとらないオプションフラグは、一つのハイフンのあとにまとめて指定できません。
誤った指定例: \$ cmd -abc
正しい指定例: \$ cmd -a -b -c
- フラグ引数を必要とするオプションフラグのフラグ引数は省略できません。
- オプションフラグとフラグ引数の間には空白またはタブが必要です。
誤った指定例: \$ cmd -afile
正しい指定例: \$ cmd -a file
- 同じオプションフラグは 2 回以上指定できません。例えば、「\$ cmd -a 1 -a 2」とは入力できません。
- ハイフンだけのオプションは入力できません。例えば、「\$ cmd -」と入力すると「-」はコマンド引数とみなされます。

入出力

(a) 入力

4. コマンドリファレンス

入力は、すべてコマンドのオプションおよび引数の並びです。

(b) 出力

コマンド処理が正常に終了したときの結果出力は、すべて標準出力に出力されます。出力される終了コードの一覧を表 4-2 に示します。

表 4-2 終了コード一覧

終了コード	意味
0	正常終了
上記以外	正常終了以外で終了

コマンド処理がエラーとなったときの出力（終了コマンドが「0」以外）は、すべて標準エラー出力に出力されます。引数エラー（終了コードが「2」）の場合は、標準エラー出力にコマンドの使用方法（USAGE）が出力されます。出力形式は次のとおりです。

出力形式

Usage:xxxx yyyy

- xxxx：コマンド名称が出力されます。
- yyyy：コマンドの指定形式が出力されます。

出力例

```
Usage:command -a -b option_arg_1 [-c] [-d option_arg_2] ...
       {-e | -f option_arg_3}
```

コマンドが実行できるファイルへのアクセス権限

システム管理者だけが、コマンドを実行できるファイルにアクセスして、コマンドを実行できます。

EDMTextIndex (テキストインデクサーの実行)

機能

テキストインデクサーを実行します。

形式

```
EDMTextIndex { -c 処理対象文書クラス名 | -f 処理対象文書クラス定義ファイルパス }
               [-l テキスト抽出処理長上限値] [-m | -o | -p FAILED] [-s]
               [-r]
```

オプション

-c または -f は、コマンドの次に指定してください。その他のオプションフラグの指定順序は任意です。

-c 処理対象文書クラス名

単一のクラスを対象とする場合に処理対象文書クラスを指定します。

-f 処理対象文書クラス定義ファイルパス

- 一つ以上のクラスを処理対象にするときに、処理対象文書クラス定義ファイルのパスを指定します。パスの指定方法については「3.2.2 文法」を参照してください。
- 処理対象文書クラス定義ファイルについては、「3.6 処理対象文書クラス定義ファイル」を参照してください。

-l テキスト抽出処理長上限値

- 環境設定ファイルの MaxOfLength エントリと同様、テキスト抽出長の上限値を指定できます。
- 実行時はオプションフラグに指定した値が優先されます。
- オプションを指定しない場合は、環境設定ファイルの値をデフォルト値とします。

-m | -o | -p FAILED

- -m を指定すると、処理対象文書クラスの中で更新された文書を処理対象にします。
- -o を指定すると、処理対象文書クラスのすべての文書を処理対象にします。
- -p FAILED を指定すると、処理対象文書クラスの中でテキスト抽出に失敗した文書を処理対象にします。
- -m, -o, または -p オプションフラグを一緒に指定した場合はエラーメッセージ KMBV12002-E を出力してプログラムを終了します。

-s

マスタレンディションおよびサブレンディションの文書を RenditionType 検索の対象とします。

4. コマンドリファレンス

-r

テキスト抽出処理でエラーが発生したときに、エラーメッセージを出力してコマンドを終了します。

注意事項

- 同じオプションフラグを 2 回以上指定した場合、または指定できないオプションを指定した場合は、エラーメッセージ KMBV12002-E を出力してプログラムを終了します。
- フラグ引数が正しくない場合は、エラーメッセージ KMBV12003-E を出力します。

指定できるオプション（オプションフラグとフラグ引数）と省略時の処理を表 4-3 に示します。

表 4-3 指定できるオプションと省略時の処理

オプションフラグ	フラグ引数	説明	省略時の処理
-c	処理対象文書クラス名	処理対象文書クラスが一つの場合に使用します。	省略できません。 -c か -f かのどちらか一つを必ず指定してください。
-f	処理対象文書クラス定義ファイルパス	処理対象文書クラスを複数個指定する場合に使用します。	省略できません。 -c か -f かのどちらか一つを必ず指定してください。
-l	テキスト抽出処理長上限値	テキスト抽出処理で出力する文書長の上限値を指定します。(単位: バイト) (指定範囲: 0 ~ 5242780)	デフォルト値は 5242780 (バイト) です。全文検索プラグインで登録できるインデクスファイルの上限値に従います。 ¹
-m	-	処理対象文書クラスの中で更新された文書を処理対象にします。	処理対象文書クラスの中で全文検索インデクスファイルが未登録の文書を処理対象にします。
-o	-	処理対象文書クラスのすべての文書を処理対象にします。	処理対象文書クラスの中で全文検索インデクスファイルが未登録の文書を処理対象にします。
-p	FAILED	処理対象文書クラスの中でテキスト抽出に失敗した文書を処理対象にします。	処理対象文書クラスの中で全文検索インデクスファイルが未登録の文書を処理対象にします。

オプションフラグ	フラグ引数	説明	省略時の処理
-r	-	テキスト抽出処理でエラーが発生した場合、文書のインデクス登録処理を中断し、コマンドを終了します。このとき、エラーとなった文書は、テキスト抽出に失敗した文書として扱います。 2, 3	ワーニングメッセージを出力して、文書のインデクス登録処理を続行します。
-s	-	マスタレンディションに RenditionType 検索の対象となる文書がない場合、サブレンディションの文書を検索の対象にします。	マスタレンディションの文書だけを RenditionType 検索の対象にします。

注 1

上限値を超えるデータが打ち切られた文書の場合、メッセージ KMBV12009-I が出力されます。ただし、XML 文書の場合は、KMBV12009-I の出力後に KMBV12037-W が出力されます。

注 2

テキスト抽出処理でエラーが発生したときのワーニングメッセージと、-r オプション指定時のエラーメッセージの対応を次に示します。

ワーニングメッセージ	エラーメッセージ	エラー内容	-r オプション指定時の動作
KMBV12035-W	KMBV12301-E	構造指定全文検索 ファイル作成時にエラー	文書のインデクスの登録処理を中断し、エラーメッセージを出力してコマンドを終了します。
KMBV12036-W	KMBV12302-E	構造指定全文検索 ファイル作成時にエラー	
KMBV12037-W	KMBV12303-E	文書のテキストデータ登録時にエラー	
KMBV12044-W	KMBV12305-E	文書のテキスト抽出処理が失敗	
KMBV12048-W	KMBV12306-E	文書のテキスト抽出処理が失敗	
KMBV12251-W	KMBV12307-E	文書のテキスト抽出処理が失敗	
KMBV12252-W	KMBV12308-E	文書のテキスト抽出処理が失敗	
KMBV12254-W	KMBV12309-E	プロパティ値の取得処理が失敗	

注 3

-r オプション指定時にテキスト抽出処理でエラーが発生した場合は、書誌情報の登録指定は無視されます。文書のインデクス登録処理を中断し、エラーメッセージを出力してからコマンドを終了します。

4. コマンドリファレンス

コマンド実行例

(指定例 1)

```
EDMTextIndex -c InputDocumentClass
```

処理対象文書クラス InputDocumentClass に所属する文書の中で、全文検索インデクスファイルが未登録の文書に対して処理を実行します。

(指定例 2)

```
EDMTextIndex -c InputDocumentClass -o
```

処理対象文書クラス InputDocumentClass に所属するすべての文書に対して処理を実行します。全文検索インデクスファイルが登録済みの文書も処理対象になり、全文検索インデクスファイルを上書きします。

(指定例 3)

```
EDMTextIndex -c InputDocumentClass -m
```

処理対象文書クラス InputDocumentClass に所属する文書の中で、更新された文書に対して処理を実行します。

(指定例 4)

```
EDMTextIndex -c InputDocumentClass -l 4000
```

処理対象文書クラス InputDocumentClass に所属する文書の中で、全文検索インデクスファイルが未登録の文書に対して処理を実行します。

1 文書当たりのテキスト抽出処理文書長の上限値を 4,000 バイト (MaxOfLength) とします。

(指定例 5)

```
EDMTextIndex -c inputDocumentClass -p FAILED
```

処理対象文書クラス InputDocumentClass に所属するテキスト抽出が失敗した文書に対して処理を実行します。

(指定例 6)

```
EDMTextIndex -c inputDocumentClass -r
```

処理対象文書クラス InputDocumentClass に所属する、すべての未登録文書に対して処理を実行します。テキスト抽出処理でエラーが発生した場合、文書のインデクス登録処理を中断し、コマンドを終了します。エラーとなった文書は、テキスト抽出に失敗した文書として扱います。

EDMCrtTIDF (テキストインデクサー定義ファイル生成)

機能

テキストインデクサーを実行するために必要な定義ファイルを自動生成します。

形式

EDMCrtTIDF ユーザID [-d 定義ファイル名]

オプション

ユーザ ID

HiRDB にテーブルを作成した認可識別子を指定します。

-d 定義ファイル名

生成する定義ファイルのパスを指定します。パスの指定方法については「3.2.2 文法」を参照してください。

このオプションを省略した場合は、「.¥EDM_Tid_define.txt」が仮定されます。

注意事項

- このコマンドを実行できるユーザは、DocumentBroker Server のシステム管理者です。
- このコマンドを実行する前に DocumentBroker Server の動作環境が整っている必要があります。
- このコマンドを実行するときは、HiRDB を起動しておく必要があります。
- このコマンドは、EDMInitMeta コマンドを実行して HiRDB に登録されたメタ情報を参照し、メタ情報に定義されているユーザクラスとユーザプロパティをすべて出力します。したがって、テキストインデクサー実行時に必要な処理対象文書クラスおよび書誌情報プロパティ以外のクラスおよびプロパティも出力されます。

5

プラグイン関数

この章では、テキストインデクサーが定義するプラグイン関数について説明します。

テキストインデクサーが定義するプラグイン関数

プラグイン関数の処理

説明する項目

フィルタプラグイン関数 (`dbrIdxTextFilter()`)

プロパティ取得関数 (任意の名称)

テキストインデクサーが定義するプラグイン関数

テキストインデクサーは、次のプラグイン関数を定義しています。

- テキストインデクサーが定義するフィルタ呼び出しインタフェースとして、テキスト抽出するためのプラグイン関数
- テキストインデクサーが定義するプロパティ値取得インタフェースとして、プロパティ値を取得するためのプラグイン関数

プラグイン関数一覧

テキストインデクサーの定義するプラグイン関数を次に示します。

関数名	機能概要
dbrIdxTextFilter()	テキスト抽出する対象ファイルのテキストを抽出します。
XXXXXXXXXX()	プラグイン環境設定ファイル (PlugInconfig.ini) の function セクションの Argv エントリに指定したプロパティ値を取得します。

ヘッダファイルについて

テキスト抽出フィルタプラグインライブラリ、およびプロパティ値取得用プラグインライブラリを作成する場合、必ず次のヘッダファイルを使用してください。

ファイル名	内容
EDM_Idx_com.h	定義するプラグイン関数のプロトタイプを宣言するヘッダファイル

コンパイルオプションについて

フィルタプラグイン機能を使用する場合は、Visual C++5.0 Service Pack 3 以降を使用してコンパイルしてください。

```

/MD
ランタイムライブラリの利用 [ マルチスレッドDLL ]
/Gd
呼び出し規則 [ _cdecl ]
/GX
例外処理の有効化
/D
次のシンボルを設定してください。
WIN32

```

プラグイン関数の呼び出し規約について

プラグイン関数は、必ず次の呼び出し規約に従って作成してください。

関数名	呼び出し規約
dbrIdxTextFilter()	__stdcall
XXXXXXXXXX()	__stdcall

プラグイン関数の処理

ここでは、テキストインデクサーによるプラグイン関数の処理を説明します。説明では、処理対象の文書が3件あり、RenditionType がそれぞれ application/rtf、application/msword、application/pdf である場合を例とします。

RenditionType 定義ファイルの定義内容

```
application/rtf/:filterE_0001
application/msword:filterE_0002
application/pdf:filterB
```

プラグイン環境設定ファイルの定義内容

```
[filterE_0001]
Library=C:¥IndexLoader¥libfilter_1.dll
[filterE_0002]
Library=C:¥IndexLoader¥libfilter_2.dll
```

```
[function]
RenditionType=application/rtf
Library=C:¥IndexLoader¥libfilter_1.dll
Function=usrGertProp1
Argv=usrProp_1
[function]
RenditionType=application/tif
Library=C:¥IndexLoader¥libfilter_2.dll
Function=usrGertProp2
Argv=usrProp1
Argv=usrProp2
```

テキストインデクサーの処理内容

テキストインデクサー（EDMTextIndex コマンド）を実行すると、テキストインデクサーは次の順序で処理を実行します。

1. RenditionType が application/rtf である文書のインデクス登録処理を実行します。
 - テキストインデクサーがテキスト抽出プラグインライブラリ（/IndexLoader/libfilter_1.sl）を呼び出し、文書からテキストを抽出します。
 - 抽出したテキストをデータベースに登録します。
 - テキストインデクサーが引数に usrProp1 の値を指定して、プロパティ値取得用プラグインライブラリ（/IndexLoader/libfilter_1.sl の usrGetProp1()）を呼び出します。
2. RenditionType が application/msword である文書のインデクス登録処理を実行します。
 - テキストインデクサーがテキスト抽出プラグインライブラリ（/IndexLoader/libfilter_2.sl）を呼び出し、文書からテキストを抽出します。
 - 抽出したテキストをデータベースに登録します。
3. RenditionType が application/tif である文書のインデクス登録処理を実行します。
 - テキストインデクサーが DocumentFilter for Text Search のテキスト抽出ライブラリを呼び出し、文書からテキストを抽出します。
 - 抽出したテキストをデータベースに登録します。

- テキストインデクサーが引数に `usrProp1` , `usrProp2` の値を指定して、プロパティ値取得用プラグインライブラリ (`/IndexLoader/libfilter_2.sl` の `usrGetProp2()`) を呼び出します。

説明する項目

各関数で説明する項目を次に示します。

機能

関数の機能について説明します。

形式

関数の記述形式を示します。

引数

関数の引数を示します。後ろに《入力》が付いている引数は、テキストインデクサーが値を設定することを示します。後ろに《出力》が付いている引数は、ユーザが作成した関数が値を設定することを示します。

戻り値

関数の戻り値を示します。

注意事項

関数を使用する上での注意事項を示します。

構造体の定義

構造体の定義内容を示します。

フィルタプラグイン関数 (dbrIdxTextFilter())

機能

全文検索用のテキストデータを抽出します。

形式

```
int    dbrIdxTextFilter (
        Idx_DocumentInfo *DocumentInfo,
        unsigned long    DataMode,
        char              *TargetData,
        unsigned long    TargetLength,
        unsigned long    MaxTextLength,
        char              *TextData,
        unsigned long    *TextLength,
        unsigned long    *OutFileDeleteFlag);
```

引数

DocumentInfo 《入力》

Idx_DocumentInfo 構造体を指定します。Idx_DocumentInfo 構造体には、処理対象文書オブジェクトのシステムプロパティ情報を指定します。

DataMode 《入力》

プラグイン環境設定ファイルの DataMode エントリに MEMORY を指定した場合は 0 を、FILE を指定した場合は 1 を指定します。

TargetData 《入力》

DataMode に 0 を指定した場合

テキスト抽出する対象データのアドレスを指定します。対象データは、テキストインデクサーがデータベースから取り出し、TargetData の先頭アドレスから対象データ長 (TargetLength) 分だけコピーします。

DataMode に 1 を指定した場合

テキスト抽出する対象ファイルのフルパスを、最後が ¥0 で終わる文字列で指定します。テキストインデクサーは対象ファイルをデータベースから取り出して環境変数 TMPDIR 下に作成し、この関数の呼び出し後に対象ファイルを削除します。

TargetLength 《入力》

テキスト抽出する対象データのデータ長を指定します。

MaxTextLength 《入力》

テキスト抽出する上限値を指定します。この値は、環境設定ファイルの MaxOfLength エントリ、または EDMTextIndex コマンドの -l オプションのどちらかで指定します。どちらにも値を指定した場合は、-l オプションに指定した値が優

5. プラグイン関数

先されます。

TextData 《出力》

DataMode に 0 を指定した場合

テキストインデクサーが MaxTextLength で指定した値（単位：バイト）だけ確保した領域に、テキスト抽出後のデータを格納します。

DataMode に 1 を指定した場合

テキストインデクサーが確保した領域（256 バイト）に、テキスト抽出後のファイルのフルパスを最後が ¥0 で終わる文字列で指定してください。この場合、EDMTextIndex コマンドを実行したユーザに、テキスト抽出後のファイルの読み取り権限が必要です。

TextLength 《出力》

DataMode に 0 を指定した場合

テキスト抽出後のデータのデータ長を指定してください。

DataMode に 1 を指定した場合

この指定は無視されます。

OutFileDeleteFlag 《出力》

DataMode に 0 を指定した場合

この指定は無視されます。

DataMode に 1 を指定した場合

TextData に指定したテキスト抽出後のファイルを削除する場合は 1 を指定してください。削除しない場合は 0 を指定してください。

戻り値

戻り値	内容	テキストインデクサーの対処
0	正常終了	処理を継続します。
正の値	テキスト抽出に失敗しました。この値はユーザが定義します。	ワーニングメッセージ KMBV12252-W を出力し、文書のインデクス登録処理を継続します。書誌情報の登録を指定している場合は、書誌情報の登録処理を実行します。書誌情報を登録しないように指定している場合は、エラーとなった文書はテキスト抽出に失敗した文書として扱います。
負の値	テキスト抽出に失敗しました。この値はユーザが定義します。	エラーメッセージ KMBV12308-E を出力し、文書のインデクス登録処理を中断し、コマンドを終了します。エラーとなった文書は、テキスト抽出に失敗した文書として扱います。

注意事項

戻り値に正の値を設定しても、EDMTextIndex コマンドで -r オプションを指定した場合は、エラーメッセージ（KMBV12308-E）を出力して文書のインデクス登録処理

を中断し、コマンドを終了します。

この関数内では throw しないでください。throw した場合は、エラーメッセージ (KMBV12308-E) を出力して文書のインデクス登録処理を中断し、コマンドを終了します。

構造体の定義

```
struct IDX_DOCUMENTINFO {
    char *pidxDVSubOID; //処理対象文書クラスのOID
    char *pidxRenditionOID; //処理対象文書 (レンディション) のOID
    char *pidxRenditionType; //処理対象文書 (レンディション) の
    RenditionType
    char *pidxRetrievalName; //処理対象文書 (コンテンツ) のRetrievalName
} Idx_DocumentInfo;
```

pidxDVSubOID

処理対象文書クラスの OID を指定します。

pidxRenditionOID

処理対象文書 (レンディション) の OID を指定します。

pidxRenditionType

処理対象文書 (レンディション) の RenditionType を指定します。このとき、処理対象文書の RenditionType プロパティから「MIME:」を除いた形式で指定します。

pidxRetrievalName

処理対象文書 (コンテンツ) の dmaPrp_RetrievalName を指定します。

プロパティ取得関数（任意の名称）

機能

関数名 (xxxxxxxx) は、ユーザが 1 ~ 128 バイト以内で定義してください。

処理対象文書クラスのプロパティ値を取得します。プラグイン環境設定ファイル (PlugInconfig.ini) の function セクションの Argv エントリに指定したプロパティが、取得対象となります。指定できるプロパティは、データ型が String, Integer32, または Boolean である文書オブジェクトのユーザプロパティです。

この関数は、テキスト抽出フィルタプラグインライブラリによるテキスト抽出が成功し、さらに抽出したテキストのデータベースへのインデクス登録が成功した場合に呼び出されます。テキスト抽出フィルタプラグインライブラリによるテキスト抽出が失敗した場合は、書誌情報のデータベースへのインデクス登録が成功しても、この関数は呼び出されません。

形式

```
int xxxxxxxxxx (
    Idx_DocumentInfo *DocumentInfo,
    int argc,
    Idx_Prop argv[]);
```

引数

DocumentInfo 《入力》

Idx_DocumentInfo 構造体を指定します。Idx_DocumentInfo 構造体には、処理対象文書オブジェクトのシステムプロパティ情報を指定します。

argc 《入力》

argv の要素を指定します。argv の要素は、プラグイン環境設定ファイル (PlugInconfig.ini) の function セクション内に指定した Argv エントリの数です。

argv 《入力》

Idx_Prop 構造体の配列を指定します。Idx_Prop 構造体の配列の要素は、プラグイン環境設定ファイル (PlugInconfig.ini) の function セクション内に指定した Argv エントリの順番に作成します。Idx_Prop 構造体の配列の要素数は、function セクション内に指定した Argv エントリの数です。

戻り値

戻り値	内容	テキストインデクサーの対処
0	正常終了	処理を継続します。

戻り値	内容	テキストインデクサーの対処
正の値	プロパティ値の取得処理に失敗しました。この値は、ユーザが定義します。	ワーニングメッセージ KMBV12254-W を出力し、文書のインデクス登録処理を継続します。
負の値	プロパティ値の取得処理に失敗しました。この値はユーザが定義します。	エラーメッセージ KMBV12309-E を出力し、文書のインデクス登録処理を中断し、コマンドを終了します。

注意事項

戻り値に正の値を設定しても、EDMTextIndex コマンドで `-r` オプションを指定した場合は、エラーメッセージ KMBV12309-E を出力して文書のインデクス登録処理を中断し、コマンドを終了します。

この関数内では `throw` しないでください。throw した場合は、エラーメッセージ KMBV12309-E を出力して文書のインデクス登録処理を中断し、コマンドを終了します。

Argv エントリに不正なプロパティ名を指定した場合は、エラーメッセージ (KMBV12309-E および KMBV12053-E) を出力し、処理を中断します。

構造体の定義

```
struct IDX_PROP{
    char          *pidxPropName;    //プロパティ名
    unsigned long idxDataSize;     //プロパティ値の領域長
    union {
        int      *pidxBoolearn;    //Boolearn型
        int      *pidxInteger32;   //Integer32型
        char     *pidxString;      //String型
    } unvalue;
} Idx_Prop;
```

pidxPropName

プロパティ名を設定します。この値は、プラグイン環境設定ファイル (PlugInconfig.ini) の function セクションの Argv エントリに指定したプロパティ名です。

共用体 uniValue のメンバ

idxDataSize にはプロパティ値の領域長を指定します。

文書オブジェクトのユーザプロパティ値は、共用体 uniValue のメンバ

pidxBoolearn, pidxInteger32, pidxString のうち、プロパティに設定されたデータ型と同一のデータ型の領域に設定されます。

なお、該当するプロパティがデータベースの NULL 値の場合は、NULL が設定されます。

6

トラブルシューティング機能

この章では、テキストインデクサーを保守運用するためのトラブルシューティング機能とメッセージについて説明します。

6.1 トラブルシューティング機能の概要

6.2 障害処理

6.3 メッセージの形式

6.4 メッセージの詳細

6.1 トラブルシュート機能の概要

この節では、テキストインデクサーを保守運用するために必要なトラブルシュート機能の概要について説明します。

(1) メッセージ

障害対応の出力メッセージのうち、重要度の高いものはイベントビューアのアプリケーションログに出力されます。

また、コマンド関連のメッセージは、このほかに、標準出力や標準エラー出力にも出力されます。

テキストインデクサーのメッセージプリフィクスは、「KMBV」です。

(2) トレース出力

処理中に発生したメッセージがトレースログファイルに出力されます。出力するメッセージは「6.4 メッセージの詳細」に記載したメッセージのうちで出力先が < F > のメッセージだけです。

(3) エラーログファイル

EDMTextIndex コマンドの処理中に発生したメッセージをエラーログファイルに出力します。該当するメッセージは、「6.4 メッセージの詳細」に記載したメッセージのうちで出力先が < U > のメッセージだけです。

6.2 障害処理

この節では、テキストインデクサーの障害処理について説明します。

(1) 全体の処理時間の上限値

テキストインデクサーの実行時に全体の処理時間の上限値を設定できます。上限値は、環境定義ファイルに指定する環境定義パラメタ (LimitTime) で指定します。デフォルトは無制限です。

テキストインデクサー実行中に、指定した処理時間の上限値に達すると、処理を打ち切ります。それまでに処理した文書については、全文検索インデクスに登録します。

テキストインデクサーが LimitTime タイムアウトで終了したときの動作については、「2.6(11) テキストインデクサーが LimitTime タイムアウトで終了したとき」を参照してください。

(2) 処理文書長の上限値

テキストインデクサーの実行時に処理する文書の長さの上限値を設定できます。上限値は環境定義ファイルに指定する環境定義パラメタ (MaxOfLength) またはテキストインデクサーの実行コマンド (EDMTextIndex) のオプションフラグ (-l) で指定します。

上限値を指定して実行した場合は、指定した上限値の範囲内の長さの文書に対しては正常に処理され、文書長が上限値に達した時点で処理が打ち切られます。ただし、XML 文書の場合は、メッセージ KMBV12037-W が出力され、指定した上限値分の登録処理は実行されません。

(3) テキスト情報抽出処理中にエラーが発生した場合

文書からテキスト情報を抽出する処理の途中にエラーが発生した場合は、該当文書を処理失敗文書として登録して、その次の処理対象文書を処理します。処理が失敗した文書については、「3.3.4 抽出処理が失敗した文書の取り扱い」を参照してください。

文書からテキスト情報を抽出する処理の途中に重大なエラーが発生した場合は、テキストインデクサーのすべての処理を終了することがあります。

6.3 メッセージの形式

この節では、テキストインデクサーのメッセージの形式について説明します。

6.3.1 メッセージの出力形式

テキストインデクサーのメッセージの出力形式は次のとおりです。

メッセージID メッセージテキスト

6.3.2 メッセージの記述形式

(1) 記述形式

「6.4 メッセージの詳細」でのメッセージの記述形式は次のとおりです。

```
KMBVnnnnn-i <Y>
日本語のメッセージテキスト
英語のメッセージテキスト
  (必要に応じて) 出力されるメッセージテキストに対する補足説明
  (S)               テキストインデクサーの動作 (処理)
  (0)               ユーザの対処方法
```

(2) メッセージ ID の記号の説明

メッセージ ID の記号の意味を説明します。

KMBV

メッセージを出力したプログラム (テキストインデクサー) を表します。

nnnnn

メッセージの番号です。それぞれのメッセージに 5 けたの固有の番号が付いています。

i

メッセージの種類を表します。

I

システムの動作を通知します。

W

処理は続行しますが、障害が発生したので警告します。

E

障害が発生したので、処理を中断します。

<Y>

メッセージの出力先を表します。

C

イベントビューア

P

標準エラー出力

F

トレースログファイル

U

エラーログファイル

注意

この記号はメッセージ ID には付いていません。このマニュアル内で使用している追加情報です。

(3) メッセージテキストの説明

出力されるメッセージテキストを示します。なお、メッセージテキストの %n (n は挿入句の順番) などの文字は、メッセージが出力される状況によって変わる値です。

また、メッセージテキストの挿入句に文書ファイル名が表示される場合、文書空間の文字コード種別の指定内容によって文書ファイル名の表示形式が異なります。

文書空間の文字コード種別が Shift-JIS の場合

文書ファイル名 [文書オブジェクトの OIID 値]

文書空間の文字コード種別が UTF-8 の場合

[文書オブジェクトの OIID 値]

6.4 メッセージの詳細

この節では、テキストインデクサーのメッセージの詳細について説明します。

KMBV12000-I <C,P,F,U>

テキストインデクサーの処理を開始しました。

TextIndex started.

テキストインデクサーの処理を開始しました。

(S)

処理を開始しました。

(O)

-

KMBV12001-I <C,P,F,U>

テキストインデクサーの処理が終了しました。

TextIndex ended.

テキストインデクサーの処理が終了しました。

(S)

処理を終了します。

(O)

-

KMBV12002-E <P>

指定したコマンド形式に誤りがあります。コマンドの形式は次のとおりです。

```
EDMTextIndex {-c 指定文書クラス名 |-f 指定文書クラス定義ファイルのパス} [-l 処理文書長上限値] [-m|-o|-p FAILED] [-s][-r]
```

Error exists in command line, usage is as follows

```
Usage : EDMTextIndex {-c InputDocumentClass} [-f InputDocumentClassFile] [-l MaxOfLength] [-m|-o|-p FAILED] [-s][-r]
```

指定したコマンド形式に誤りがあります。

(S)

処理を終了します。

(O)

正しいコマンド形式を指定して再度実行してください。

KMBV12003-E <P>

コマンドオプション (%1) に指定した値が不正です。

値 : %2

Invalid value exists in CommandOption(%1).

value : %2

コマンドオプション (%1) に指定した値が正しくないため、処理を続行できません。

(S)

処理を終了します。

(O)

エラーの要因を取り除いてください。

KMBV12004-E <P>

エラーログファイルのオープンでエラーが発生しました。

要因コード : %1

An error occurred in opening of ErrorLog file.

Reason code : %1

エラーログファイルのオープンでエラーが発生しました。

(S)

処理を終了します。

(O)

表 6-1 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-1 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	環境設定ファイルの ErrorLog エントリに指定したパスが存在しません。	指定したパスを確認してください。
2	パスの長さが制限値 (255 バイト) を超えています。	指定したパスを確認してください。
3	ディレクトリに書き込み権限 / アクセス権限がありません。	指定したディレクトリの権限を確認してください。
4	ファイルに書き込み権限 / アクセス権限がありません。	指定したファイルの権限を確認してください。

KMBV12005-E <C,P,F>

環境変数の指定に誤りがあります。

環境変数名 : %1

Environment variable is invalid.

Environment variable name : %1

環境変数 (%1) が定義されていないか、指定値が誤っているため、コマンドが実行できません。

%1 : テキストインデクサーの実行には、\$DOCBROKERDIR,\$TEXTINDEXDIR の環境変数を正しく定義する必要があります。

(S)

処理を終了します。

(O)

環境変数を正しく定義してから再度実行してください。

KMBV12006-E <P>

テキストインデクサーが実行中のため、コマンドが実行できません。

Unable to execute this command because TextIndex is executing.

テキストインデクサーを複数実行しようとしています。

(S)

処理を終了します。

(O)

実行中の処理が終了してから再度実行してください。

KMBV12007-I <C,P,F,U>

処理対象文書クラス (%1) の処理が終了しました。

登録 : %2 件

エラー : %3 件

実行文書 : %4 件

InputClass(%1) ended.

Entry : %2

Error : %3

Execute document : %4

処理対象文書クラス (%1) のうち %4 件まで実行しました。その結果、%2 件が登録で

き、エラーによって %3 件が登録できませんでした。

(S)

処理を終了します。

(O)

-

KMBV12008-E <C,P,F,U>

テキストインデクサー実行時間上限値を超えました。

An error occurred in processing TextIndex for LimitTime Over.

テキストインデクサー実行中にテキストインデクサー実行時間上限値を超えました。

(S)

処理を終了します。

(O)

-

KMBV12009-I <F,U>

文書 (%1) 処理中にテキスト抽出で出力する文書長の上限値を超えました。

Text Search stopped in extracting text data from document(%1) for MaxOfLength Over.

処理対象文書クラスの文書 (%1) のテキスト抽出処理中に、テキスト抽出で出力する上限値を超えたためテキスト抽出処理を打ち切りました。

(S)

処理を続行します。

(O)

-

KMBV12010-E <C,P,F>

環境設定ファイル (TextIndex.ini) のオープンでエラーが発生しました。

要因コード : %1

An error occurred in opening of TextIndex.ini file.

Reason code : %1

環境設定ファイルのオープンでエラーが発生しました。要因コード (%1) でエラー内容を確認してください。

(S)

処理を終了します。

6. トラブルシュート機能

(O)

表 6-2 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-2 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	環境設定ファイルが存在しません。	<インストールディレクトリ> ¥sample¥TextIndex.ini を <インストールディレクトリ> ¥etc にコピーして回復させてください。
2	環境設定ファイルにアクセス権がありません。	環境設定ファイルのアクセス権を確認してください。
3	環境設定ファイルに定義情報がありません。	<インストールディレクトリ> ¥sample¥TextIndex.ini を <インストールディレクトリ> ¥etc にコピーして回復させてください。
4	パスの長さが制限値 (255 バイト) を超えています。	インストールし直して、環境設定ファイルのパス名が 255 バイト以下となるディレクトリに実行環境を作成し直してください。

KMBV12011-E <C,P,F,U>

RenditionType 定義ファイルのオープンでエラーが発生しました。

要因コード : %1

An error occurred in opening of RenditionType file.

Reason code : %1

RenditionType 定義ファイルのオープンでエラーが発生しました。要因コード (%1) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-3 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-3 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	RenditionType 定義ファイルが存在しません。	環境設定ファイルの RenditionType_def パラメタに記述している RenditionType 定義ファイルが存在するかどうかを確認してください。
2	RenditionType 定義ファイルにアクセス権限がありません。	RenditionType 定義ファイルのアクセス権限を確認してください。
3	RenditionType 定義ファイルに定義情報がありません。	RenditionType 定義ファイルの内容を確認してください。
4	パスの長さが制限値（255 バイト）を超えています。	環境設定ファイルの RenditionType_def パラメタに記述している RenditionType 定義ファイルのパス名が 255 バイト以下となるように変更してください。

KMBV12012-E <C,P,F,U>

環境定義パラメタの指定に誤りがあります。

環境定義パラメタ名 : %1

Environment parameter variable is invalid.

Environment parameter variable name : %1

環境定義パラメタ (%1) が定義されていないか、指定値が誤っているため、コマンドが実行できません。テキストインデクサーの実行には、%1 の環境定義パラメタを正しく定義する必要があります。

(S)

処理を終了します。

(O)

環境定義パラメタを定義してから再度実行してください。

KMBV12013-E <C,P,F>

処理対象文書クラス定義ファイル (%1) のオープンでエラーが発生しました。

要因コード : %2

An error occurred in opening of InputDocument(%1).

Reason code : %2

処理対象文書クラス (%1) のオープンでエラーが発生しました。

(S)

処理を終了します。

6. トラブルシュート機能

(O)

表 6-4 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-4 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	処理対象文書クラス定義ファイルが存在しません。	処理対象文書クラス定義ファイルが存在するかどうかを確認してください。
2	処理対象文書クラス定義ファイルにアクセス権限がありません。	処理対象文書クラス定義ファイルのアクセス権限を確認してください。
3	処理対象文書クラス定義ファイルに定義情報がありません。	処理対象文書クラス定義ファイルの内容を確認してください。
4	パスの長さが制限値 (255 バイト) を超えています。	処理対象文書クラス定義ファイルのパス名が 255 バイト以下となるように変更してください。

KMBV12015-E <C,P,F,U>

MaxOfLength エントリに指定した値が不正です。

Invalid value exists in MaxOfLength entry.

コマンドまたは環境設定ファイルの MaxOfLength エントリに指定した値が、数値ではないか、指定範囲 (0 ~ 5242780) 外の値です。

(S)

処理を終了します。

(O)

コマンドまたは環境設定ファイルの MaxOfLength エントリに指定した値を確認してください。

KMBV12016-E <C,P,F,U>

LimitTime エントリに指定した値が不正です。

Invalid value exists in LimitTime entry.

環境設定ファイルの LimitTime エントリに指定した値が、数値ではないか、指定範囲 (1 ~ 2147483647) 外の値です。

(S)

処理を終了します。

(O)

環境設定ファイルの LimitTime エントリに指定した値を確認してください。

KMBV12017-E <C,P,F,U>

ModeOfProperty エントリに指定した値が不正です。

Invalid value exists in ModeOfProperty entry.

環境設定ファイルの ModeOfProperty エントリに指定した値が不正です。

(S)

処理を終了します。

(O)

環境設定ファイルの ModeOfProperty エントリに指定した値を確認してください。

KMBV12018-E <C,P,F,U>

PropertyOfIndex エントリに指定した書誌情報が不正です。

Invalid value exists in PropertyOfIndex entry.

環境設定ファイルの PropertyOfIndex エントリに、長さの制限値 (128 バイト) を超えた文字列が指定されています。

(S)

処理を終了します。

(O)

環境設定ファイルの PropertyOfIndex エントリに指定した書誌情報を確認してください。

KMBV12020-E <C,P,F,U>

定義ファイル (%1) のオープンでエラーが発生しました。

要因コード : %2

An error occurred in opening of Define file(%1).

Reason code : %2

環境定義パラメタ (Definition_def) に指定された定義ファイル (%1) のオープンでエラーが発生しました。要因コード (%2) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-5 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-5 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	定義ファイルが存在しません。	環境設定ファイルの Definition_def パラメータに記述している定義ファイルが存在するかどうかを確認してください。
2	定義ファイルにアクセス権限がありません。	定義ファイルのアクセス権限を確認してください。
3	定義ファイルに定義情報がありません。	EDMCrtTIDF コマンドを利用して定義ファイルを再作成してください。
4	パスの長さが制限値（255 バイト）を超えています。	環境設定ファイルの Definition_def パラメータに記述している定義ファイルのパス名が 255 バイト以下となるように変更してください。

KMBV12021-E <C,P,F,U>

%1 エントリの GUID 値が不正です。

Invalid GUID value exists in %1 entry.

定義ファイルの [ClassNameDefinition] または [PropNameDefinition] に指定したエントリ名 (%1) の GUID 値が、長さが 36 文字でないか、指定できる文字以外を指定しているか、- の位置が不正です。

(S)

処理を終了します。

(O)

定義ファイルを修正してください。

KMBV12022-E <C,P,F,U>

セクションに指定したエントリ名称が不正です。

セクション名 : %1

エントリ名 : %2

Invalid entry name exists in section.

Section name : %1

Entry name : %2

定義ファイルの %1 (ClassNameDefinition セクションまたは PropNameDefinition セクション) のエントリ名称 (%2) に " * * ", "@", ".", "," または半角空白が含まれているか、長さの制限値 (128 バイト) を超えた名称が指定されています。

(S)

処理を終了します。

(O)

定義ファイルの定義を修正してください。

KMBV12023-W <F,U>

環境設定ファイルに指定したエントリ (%1) が複数存在します。

Specified entry (%1) in environment file exist more than one.

環境設定ファイルに指定したエントリ (%1) が複数存在します。最上位の指定を有効にして処理を続行します。

(S)

処理を続行します。

(O)

-

KMBV12024-W <F,U>

RenditionType 定義ファイルに指定した RenditionType (%1) が複数存在します。

Specified RenditionType (%1) in RenditionType file exist more than one.

RenditionType 定義ファイルに同じ RenditionType が存在します。RenditionType (%1) を無視して処理を続行します。

(S)

処理を続行します。

(O)

-

KMBV12025-W <F,U>

処理対象文書クラス定義ファイルに指定したクラス (%1) が複数存在します。

Specified Class (%1) in InputDocument file exist more than one.

処理対象文書クラス定義ファイルに指定したクラス (%1) が複数存在します。最上位の指定を有効にして処理を続行します。

(S)

処理を続行します。

(O)

-

KMBV12026-E <C,P,F,U>

指定された処理対象文書クラス名 (%1) はデータベースにありません。

InputClass(%1) not found in Database.

指定された処理対象文書クラス名 (%1) はデータベースにありません。

(S)

処理を終了します。

(O)

データベースの構築が正しいかを確認してください。

KMBV12027-W <F,U>

同一の書誌情報プロパティ名 (%1) が存在しています。

Property name (%1) in PropertyOfIndex entry exist more than one.

環境定義ファイルの PropertyOfIndex エントリに同一の書誌情報プロパティ名が存在しています。最上位の指定を有効にして処理を続行します。

(S)

処理を続行します。

(O)

-

KMBV12028-E <C,P,F,U>

RenditionType 定義ファイルに指定した RenditionType (%1) が不正です。

Invalid RenditionType(%1) exists in RenditionType file.

RenditionType 定義ファイルに指定した RenditionType (%1) が正しくありません。

(S)

処理を終了します。

(O)

RenditionType 定義ファイルを修正してください。

KMBV12029-E <C,P,F,U>

処理対象文書クラス定義ファイルに指定したクラス名 (%1) が不正です。

Invalid class name(%1) exists in InputDocument.

処理対象文書クラス定義ファイルに指定したクラス名 (%1) が正しくありません。

(S)

処理を終了します。

(O)

処理対象クラス定義ファイルを修正してください。

KMBV12030-E <C,P,F,U>

指定された書誌情報 (%1) は定義ファイルにありません。

Property name(%1) not found in Database.

指定された書誌情報 (%1) は定義ファイルにありません。

(S)

処理を終了します。

(O)

環境設定ファイルの PropertyOfIndex エントリに指定した書誌情報プロパティが、定義ファイルにあるか確認してください。

KMBV12031-E <C,P,F,U>

定義ファイルに指定したクラス名 (%1) が複数存在します。

Specified Class name (%1) in Define file exist more than one.

定義ファイルに指定したクラス名 (%1) が複数存在します。

(S)

処理を終了します。

(O)

定義ファイルを修正してください。

KMBV12032-E <C,P,F,U>

定義ファイルに指定したプロパティ名 (%1) が複数存在します。

Specified Property name (%1) in Define file exist more than one.

定義ファイルに指定したプロパティ名 (%1) が複数存在します。

(S)

処理を終了します。

(O)

定義ファイルを修正してください。

KMBV12033-E <C,P,F,U>

定義ファイルの %1 セクションが不正です。

Invalid %1 section exists in Define file.

定義ファイルに指定されている %1 セクションがありません。

(S)

処理を終了します。

(O)

定義ファイルを修正してください。

KMBV12034-E <C,P,F,U>

指定された処理対象クラス (%1) は定義ファイルにありません。

InputClass(%1) not found in Define file.

指定された処理対象文書クラス (%1) は定義ファイルにありません。

(S)

処理を終了します。

(O)

処理対象クラス名が定義ファイルに存在するか確認してください。

KMBV12035-W <F,U>

構造指定全文検索ファイル作成時にエラーが発生しました。

詳細情報 : %1

An error occurred making struct all sentence search file.

Information %1

構文指定全文検索ファイル作成時にエラーが発生しました。

(S)

エラーが発生したオブジェクトを無視して処理を続行します。

(O)

表 6-6 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-6 要因別の対策一覧

詳細情報 (%1) に示されるメッセージ	エラーの要因	対策
HiRDB Adapter for XML loading failed.	次のどちらかの要因が考えられます。 <ul style="list-style-type: none"> • HiRDB Adapter for XML がインストールされていない • 環境変数 PATH にパスが設定されていない 	要因に応じて次のどちらかの対策を実施してください。 <ul style="list-style-type: none"> • HiRDB Adapter for XML をインストールする • 環境変数 PATH のパスを確認する

KMBV12036-W <F,U>

構造指定全文検索ファイル作成時にエラーが発生しました。

詳細情報 : %1

An error occurred making struct all sentence search file.

Information %1

構文指定全文検索ファイル作成時にエラーが発生しました。

(S)

エラーが発生したオブジェクトを無視して処理を続行します。

(O)

表 6-7 に示す対策一覧を基にエラーの要因を取り除いてください。

表 6-7 要因別の対策一覧

詳細情報 (%1) に示されるメッセージ	エラーの要因	対策
Content File writing error.	環境変数 TMPDIR 下の文書実体一時保管用ファイルの作成に失敗しました。	環境変数 TMPDIR のアクセス権限、ディスク容量を確認してください。
Preprocessing Library for Text Search loading failed.	Preprocessing Library for Text Search がインストールされていません。	Preprocessing Library for Text Search をインストールしてください。
TmpDir access error.	環境変数 TMPDIR 下の一時ファイルの読み込みに失敗しました。	環境変数 TMPDIR のアクセス権限を確認してください。
TmpFile create error.	環境変数 TMPDIR 下の全文検索インデクスー時保管用ファイルの作成に失敗しました。	環境変数 TMPDIR のアクセス権限、ディスク容量を確認してください。
XMS file writing error.	環境変数 TMPDIR 下に XMS ファイルの作成に失敗しました。	環境変数 TMPDIR のアクセス権限、ディスク容量を確認してください。
xx	全文検索インデクスファイルの作成処理に失敗しました。	xx に示されるメッセージに従って、エラーの要因を取り除いてください。

6. トラブルシュート機能

注 Preprocessing Library for Text Search から返却されたエラーメッセージテキスト

KMBV12037-W <F,U>

文書 (%1) のテキストデータ登録時にエラーが発生しました。

要因コード : %2

テキストデータ長 : %3

Error occurred in entry text data from document(%1).

Reason code: %2

Length of text data: %3

文書 (%1) のテキストデータ登録時にエラーが発生しました。テキストデータ長は %3 バイトです。

(S)

エラーが発生したオブジェクトを無視して処理を続行します。

(O)

表 6-8 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-8 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	XML 文書のテキストデータのデータ長が MaxOfLength 値を超えたため、テキストデータは登録できませんでした。	MaxOfLength 値をテキストデータ長以上に設定して再度実行してください。

KMBV12040-W <F,U>

指定した処理対象文書クラス (%1) は処理対象外のクラスです。

要因コード : %2

Invalid Class in InputClass(%1).

Reason code : %2

指定した処理対象文書クラス (%1) は処理対象外です。

(S)

処理を続行します。

(O)

表 6-9 に示す対策一覧を基にエラーの要因を取り除いてください。

表 6-9 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	<ul style="list-style-type: none"> 全文検索の対象でないクラスが指定されています。 指定したクラスに、処理対象外の全文検索インデクス用プロパティが定義されています。 	メタ情報およびデータベース定義を正しく設定してください。

KMBV12042-W <F,U>

文書 (%1) に対するテキストインデクサーの処理が終了しました。

テキスト抽出エラーにより書誌情報 (%2) をインデックスファイルに登録しました。

登録文書長 : %3

Document (%1) was executed.

Error, Property (%2) are registered as index file.

Length of index file : %3

処理対象文書クラスの文書 %1 を処理しました。その結果、文書 %1 を登録し、文書中のテキスト抽出エラーによって文書の書誌情報 %2 をインデックスファイルに登録しました。インデックスファイルの文書長は %3 です。

%2 : テキスト抽出失敗時に、代わりに登録される書誌情報として指定されたタイトル、コメント、著者など (ON 設定時)

(S)

処理を続行します。

(O)

-

KMBV12043-W <F,U>

処理対象文書クラス (%1) に、書誌情報 (%2) はデータベースにありません。

Property(%2) not found in Database of InputClass(%1).

処理対象文書クラス (%1) に、書誌情報 (%2) はデータベースにありません。

(S)

エラーが発生した処理対象文書クラスを無視して処理を続行します。

(O)

環境設定ファイルの PropertyOfIndex エントリに指定した書誌情報が、処理対象文書クラスに含まれているか確認してください。

KMBV12044-W <F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

要因コード : %2

An error occurred in extracting text data from document (%1).

Reason code : %2

文書 %1 のテキスト抽出処理中に、テキスト抽出ライブラリでエラーが発生したためテキスト抽出処理が失敗しました。

(S)

処理を続行します。

(O)

要因コード (%2) を基に、Document Filter for Text Search のドキュメントを参照し、対策を実施してください。

KMBV12045-W <F,U>

文書 (%1) は RenditionType が不正のため処理対象外の文書です。

Invalid Document (%1) for RenditionType error.

文書 %1 は、RenditionType が RenditionType 定義ファイルに定義されていません。または現在サポートされていないフィルタ拡張子のため処理対象外の文書です。

(S)

処理を続行します。

(O)

RenditionType 定義ファイルの内容を確認してください。

KMBV12046-W <F,U>

処理対象文書クラス (%1) の書誌情報 (%2) は指定できないデータ型です。

Invalid Property (%2) in InputClass (%1).

処理対象文書クラス (%1) の書誌情報 (%2) は指定できないデータ型です。指定できるデータ型は、INTEGER 型、CHAR 型、MVARCHAR 型の 3 種類です。

(S)

処理を続行します。

(O)

データベースの定義内容を確認してください。

KMBV12047-I <F,U>

文書 (%1) に対するテキストインデクサーの処理が終了しました。

Document (%1) was executed.

処理対象文書クラスの文書 %1 を処理しました。その結果、文書 %1 を登録しました。

(S)

処理を続行します。

(O)

-

KMBV12048-W <F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

エラー情報 : %2

An error occurred in extracting text data from document (%1).

Error information : %2

文書 %1 のテキスト抽出処理中に、テキスト抽出ライブラリでエラーが発生したか、TMPDIR 下の PDF 用一時ファイルの作成が失敗したため、テキスト抽出処理が失敗しました。

%2 : エラー情報

エラー情報でエラー内容を確認してください。

(S)

処理を続行します。

(O)

表 6-10 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-10 エラー情報別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
Not PDF file	PDF ファイル以外の文書が指定されています。	RenditionType 定義ファイルのフィルタ拡張子の定義が誤っていないか、Rendition の定義が誤っていないか確認してください。
PDF file Library error	PDF テキスト抽出ライブラリでエラーが発生しました。	トレースログとエラーが発生した文書を採取して保守員に連絡してください。
PDF file writing error.	PDF ファイルの書き込みに失敗しました。	環境変数 TMPDIR のアクセス権限またはディスク容量を確認してください。

6. トラブルシュート機能

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
TmpDir access error.	環境変数 TMPDIR 下の一時ファイルの読み込みに失敗しました。	環境変数 TMPDIR のアクセス権限を確認してください。
Unmatch PDF file	次のどちらかの要因によってテキスト抽出に失敗しました。 <ul style="list-style-type: none">• 処理対象の PDF ファイルのバージョンがサポート対象外である• 処理対象の PDF ファイルの書式が不正	PDF ファイルのバージョンおよび書式を確認してください。

KMBV12049-I <F,U>

文書 (%1) に対するテキストインデクサーの処理を開始しました。

Document (%1) is start.

処理対象文書クラスの文書 %1 の処理を開始しました。

(S)

処理を続行します。

(O)

-

KMBV12050-E <C,P,F,U>

データベースへの接続でエラーが発生しました。

An error occurred in connecting database.

次のどれかの要因で、データベースに接続できません。

- データベースサーバが起動していない
- DocumentBroker サーバの接続情報の定義を使用してデータベースに接続する場合に、DocumentSpace 構成定義ファイル (docspace.ini) が存在しないか、またはアクセス権がない
- 定義内容に誤りがある

(S)

処理を終了します。

(O)

データベースの起動または接続情報の定義を確認してください。

KMBV12051-E <C,P,F,U>

OIID の取得でエラーが発生しました。

An error occurred in getting OIID from database.

データベースから OIID が取得できませんでした。データベースサーバが起動していないか、データベースが壊れている可能性があります。

(S)

処理を終了します。

(O)

データベースサーバの起動またはデータベースを確認してください。

KMBV12052-E <C,P,F,U>

書誌情報の制限のため処理が続行できません。

Unable to continue processing due to a limit of Property.

書誌情報の制限のため処理が続行できません。書誌情報取得のための領域が制限値 (40,000 バイト未満) を超えたか、有効な書誌情報のプロパティ数が制限数 (4,000 個) を超えました。

(S)

処理を終了します。

(O)

有効な書誌情報の指定を制限内にしてください。

KMBV12053-E <C,P,F,U>

データベースでエラーが発生しました。

エラー情報 : %1

要因コード : %2

An error occurred in database.

Error information : %1

Reason code : %2

データベースへのアクセスでエラーが発生しました。

%1 : データベースのエラーメッセージ

%2 : データベースのエラーコード

エラー情報に出力されるメッセージ ID でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-11 に示すエラーメッセージを参照して、エラーの要因を取り除いてください。

表 6-11 %1 に示されるエラーメッセージ

メッセージ ID	確認する内容
KFPA11561-E	次の定義ファイルの内容を見直してください。 <ul style="list-style-type: none"> DocumentBroker サーバの接続情報の定義を使用してデータベースに接続する場合、DocumentSpace 構成定義ファイル (docspace.ini) の記述に誤りがあります。 テキストインデクサーの接続情報の定義を使用してデータベースに接続する場合、TextIndex.ini の記述に誤りがあります。
その他	マニュアル「HiRDB メッセージ」を参照してください。

KMBV12079-W <F,U>

文書 (%1) 処理中に文書ファイルが更新されたため、処理を中断しました。

Text Search canceled indexing document(%1), for updated content file while indexing.

処理対象文書クラスの文書 (%1) の処理中に、DocumentBroker クライアントによる文書ファイルの更新を検知したため、処理結果を登録しませんでした。

(S)

処理を続行します。

(O)

テキストインデクサーを再度実行してください。

KMBV12080-E <C,P,F,U>

メモリが不足しているため、処理を続行できません。

There is insufficient memory.

メモリが不足しました。

(S)

処理を終了します。

(O)

不要なプログラムを終了させて再度実行してください。

KMBV12081-E <C,P,F,U>

テキストインデクサーで内部矛盾が発生しました。

エラー情報 : %1

Internal error occurred to TextIndex.

Error information : %1

テキストインデクサーで内部矛盾が発生しました。

%1 : エラー情報。

(S)

処理を終了します。

(O)

トレースログを採取して保守員に連絡してください。

KMBV12082-E <C,F>

システムコールでエラーが発生しました。

関数名 : %1

要因コード : %2

An error occurred on the system call function.

Function Name : %1

Reason Code : %2

システムコール (%1) でエラーが発生しました。

%1 : エラーの発生した関数名

%2 : システムコールで返された errno 値

(S)

処理を終了します。

(O)

%2 に出力される要因コードを基に障害の要因を取り除いてください。

KMBV12083-E <C,P,F,U>

ディスク容量が不足しているため、処理を続行できません。

There is insufficient space of disk.

ディスク容量が不足しました。

(S)

処理を終了します。

(O)

ディスクに空きがあることを確認してから再度実行してください。

KMBV12084-E <C,F>

ファイルシステムに入出力エラーが発生しました。

対象 : %1

6. トラブルシュート機能

操作 : %2

要因コード : %3

付加情報 : %4

An I/O error occurred in the file system.

Object : %1

Operation : %2

Reason Code : %3

Additional Information : %4

ファイルシステム中のファイルに操作 (%2) を実行する際にエラーが発生したため、処理が続きません。ファイル (%1) へのアクセス権限がないか、または、テキストインデクサーが複数実行されている可能性があります。

(S)

処理を終了します。

(O)

DocumentBroker 管理者権限で再実行してください。テキストインデクサーが複数実行されている場合は、実行中のテキストインデクサー終了後に再度実行してください。

上記の対処を実行しても再度エラーが発生する場合は、保守員に連絡してください。

KMBV12086-E <P>

言語の指定に誤りがあります。

言語名 : %1

An error occurred on the setting of locale.

Locale name : %1

(S)

コマンドを実行した環境下の言語の指定が不正であるため、処理を中断しました。

(O)

ロケール (国または地域) に日本語を設定してください。

KMBV12087-E <C,P,F,U>

smgrrg.ini ファイルのオープンでエラーが発生しました。

要因コード : %1

An error occurred in opening of smgrrg.ini file.

Reason code : %1

DocumentBroker サーバが提供する smgrrreg.ini ファイルのオープンでエラーが発生しました。

(S)

処理を終了します。

(O)

表 6-12 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-12 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	smgrrreg.ini ファイルが存在しません。	smgrrreg.ini ファイルが存在するかどうか、次の場所を確認してください。 <ul style="list-style-type: none"> ・ < DocumentBroker サーバのインストールディレクトリ > %etc 下
2	smgrrreg.ini ファイルにアクセス権限がありません。	smgrrreg.ini ファイルのアクセス権限を確認してください。
3	smgrrreg.ini ファイルから DmaID が取得できません。	次の対策を実施してください。 <ul style="list-style-type: none"> ・ < DocumentBroker サーバのインストールディレクトリ > %adm%etc_org%smgrrreg.ini を < DocumentBroker サーバのインストールディレクトリ > %etc 下にコピーして回復させてください。
4	パスの長さが制限値 (255 バイト) を超えています。	次のファイルが 255 バイト以下になるように、DocumentBroker サーバの環境を見直してください。 <ul style="list-style-type: none"> ・ < DocumentBroker サーバのインストールディレクトリ > %etc%smgrrreg.ini ファイル

KMBV12088-E <C,P,F,U>

メタ情報の取得でエラーが発生しました。

GUID 値 : %1

An error occurred in obtaining of Meta information.

GUID value : %1

(S)

処理を終了します。

6. トラブルシュート機能

(O)

GUID 値 (%1) のメタ情報が正しく登録されているかを確認してください。

KMBV12089-W <F,U>

文書 (%1) は処理対象外の文書です。

処理対象文書クラス : %2

文書登録種別 : %3

RenditionType : %4

フィルタ拡張子 : %5

Invalid Document(%1).

InputClass : %2

Input Type : %3

RenditionType : %4

Filter : %5

処理対象文書クラス (%2) の SGMLTEXT 型列定義のプラグインのテキストデータ登録種別は (%3) です。

文書 (%1) は、RenditionType 定義ファイルに指定されている RenditionType (%4) のフィルタ拡張子 (%5) を使用するように定義されているため、テキスト抽出できません。

(S)

処理を続行します。

(O)

メタ情報およびデータベースの SGMLTEXT 型列定義を正しく設定するか、またはフィルタ拡張子を正しく指定してください。

KMBV12090-I <C,P,F>

テキストインデクサー実行環境を作成中です。

It is creating the environment of execute TextIndex.

(S)

テキストインデクサー実行環境を作成しています。

(O)

-

KMBV12091-I <C,P,F>

テキストインデクサー実行環境を作成しました。

It is success to create the environment of execute TextIndex.

(S)

テキストインデクサー実行環境の作成が完了しました。

(O)

-

KMBV12092-I <C,P,F>

テキストインデクサー実行環境を削除中です。

It is deleting the environment of execute TextIndex.

(S)

テキストインデクサー実行環境の削除中です。

(O)

-

KMBV12093-I <C,P,F>

テキストインデクサー実行環境を削除しました。

It is success to delete the environment of execute TextIndex.

(S)

テキストインデクサー実行環境の削除を完了しました。

(O)

-

KMBV12094-W <C,P,F>

ファイル(%1)は既に実行環境に存在するため、処理をスキップします。

The operation skipped about file (%1), because it already exists.

(S)

%1 に示したファイルはすでに指定した実行環境に存在しているので処理をスキップしました。

(O)

-

KMBV12095-E <C,P,F>

ファイル (%1) が既に実行環境に存在するため、処理が続行できません。

The operation can not continue, because the file (%1) already exists.

(S)

%1 に示したディレクトリを作成しようとしたますが、すでにファイルとして存在したため処理を中断しました。

(O)

ファイルを削除してください。または別の実行環境を指定してコマンドを再実行してください。

KMBV12096-E <C,P,F>

ファイル (%1) がインストールディレクトリ下に存在するため、処理が続行できません。

The operation can not continue, because the file (%1) exists under install directory.

(S)

インストールディレクトリ下に存在するファイル %1 の open システムコールに失敗したため、処理を中断しました。

(O)

コンソールまたはトレースファイルへ出力された open システムコールのエラーメッセージを確認して、エラーの要因を取り除いて再実行してください。

KMBV12097-E <C,P,F>

他のプログラムが起動されています。

Other program was started.

(S)

指定したディレクトリで、ほかのプログラムが起動しています。

(O)

-

KMBV12098-E <C,P,F>

環境作成先のディレクトリもしくはインストールディレクトリが存在しません。

位置 : %1

Directory for making environment or install directory is not exist.

Position : %1

環境作成先ディレクトリまたはインストールディレクトリが存在しません。

(S)

テキストインデクサー実行環境作成を中止します。

(O)

%1に必要なディレクトリが存在するかどうか確認してください。

KMBV12099-E <C,P,F>

環境作成先ディレクトリにアクセス権がありません。

位置 : %1

You don't have permission of directory for making environment.

Position : %1

環境作成先ディレクトリにアクセス権がありません

(S)

テキストインデクサー実行環境作成を中止します。

(O)

%1のディレクトリにアクセス権があるかどうか確認してください。

KMBV12100-E <C,P,F>

インストールディレクトリ下を指定することはできません。

Can't designate install directory.

インストールディレクトリ下へ環境を作成したり、インストールディレクトリ下の環境を削除したりすることはできません。

(S)

テキストインデクサー実行環境作成を中止します。

(O)

指定したディレクトリがインストールディレクトリ下であることを確認してください。

KMBV12202-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) のオープンでエラーが発生しました。

要因コード : %1

An error occurred in opening of PlugInconfig.ini file.

Reason code : %1

プラグイン環境設定ファイル (PlugInconfig.ini) のオープンに失敗しました。要因コード (%1) でエラーの内容を確認してください。

6. トラブルシュート機能

(S)

処理を終了します。

(O)

表 6-13 に示す対策一覧を基にエラーの要因を取り除いてから、再度実行してください。

表 6-13 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	プラグイン環境設定ファイルがありません。	<p><インストールディレクトリ> ¥sample¥PlugInconfig.ini を下記にコピーしてください。</p> <ul style="list-style-type: none"> • <インストールディレクトリ>¥etc¥PlugInconfig.ini
2	プラグイン環境設定ファイルにアクセス権がありません。	プラグイン環境設定ファイルのアクセス権を確認してください。
3	プラグイン環境設定ファイルに定義情報がありません。	<p><インストールディレクトリ> ¥sample¥PlugInconfig.ini を下記にコピーしてください。</p> <ul style="list-style-type: none"> • <インストールディレクトリ>¥etc¥PlugInconfig.ini
4	パスの長さが制限値 (255 バイト) を超えています。	<p>次のパス名が 255 バイト以内になるように、実行環境を再作成してください。</p> <ul style="list-style-type: none"> • <インストールディレクトリ>¥etc¥PlugInconfig.ini

KMBV12203-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) のセクションが不正です。

行: %1

Invalid section of PlugInconfig.ini file.

Line: %1

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 行目のセクションに指定した値が不正です。

%1: 行番号

(S)

処理を終了します。

(O)

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 行目のセクションに指定した値を確認してください。

KMBV12204-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) のエントリが不正です。

エントリ名 : %1

行 : %2

Invalid entry of PlugInconfig.ini file.

Entry name : %1

Line : %2

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 エントリが不正です。

%1 : エントリ名

%2 : 行番号

(S)

処理を終了します。

(O)

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 エントリを確認してください。

KMBV12205-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) のエントリに指定した値が不正です。

エントリ名 : %1

行 : %2

Invalid value exists in %1 entry of PlugInconfig.ini file.

Entry name : %1

Line : %2

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 エントリに指定した値が不正です。

%1 : エントリ名

%2 : 行番号

(S)

処理を終了します。

(O)

プラグイン環境設定ファイル (PlugInconfig.ini) の %1 エントリに指定した値を確認してください。

KMBV12206-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) にフィルタ拡張子セクションが定義されていません。

セクション名 : %1

Filter section of PlugInconfig.ini file is not defined.

Section name : %1

プラグイン環境設定ファイル (PlugInconfig.ini) のフィルタ拡張子セクション (%1) が定義されていません。

%1 : セクション名

(S)

処理を終了します。

(O)

プラグイン環境設定ファイル (PlugInconfig.ini) のセクションに、フィルタ拡張子セクション (%1) を指定してください。

KMBV12207-W <F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) に指定したエントリが複数存在します。

エントリ名 : %1

行 : %2

Specified entry in PlugInconfig.ini file exist more than one.

Entry name : %1

Line : %2

プラグイン環境設定ファイル (PlugInconfig.ini) に指定した %1 エントリが複数存在します。最上位の指定を有効にして処理を続行します。

%1 : エントリ名

%2 : 行番号

(S)

処理を続行します。

(O)

-

KMBV12208-W <F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) に指定したセクションが複数存在します。

セクション名 : %1

行 : %2

Specified section in PlugInconfig.ini file exist more than one.

Section name : %1

Line : %2

プラグイン環境設定ファイル (PlugInconfig.ini) に指定した %1 セクションが複数存在します。最上位の指定を有効にして処理を続行します。

%1 : セクション名

%2 : 行番号

(S)

処理を続行します。

(O)

-

KMBV12209-E <C,P,F,U>

プラグイン環境設定ファイル (PlugInconfig.ini) のセクションにエントリが定義されていません。

セクション名 : %1

エントリ名 : %2

The Section of PlugInconfig.ini file does not define an entry.

Section name : %1

Entry name : %2

プラグイン環境設定ファイル (PlugInconfig.ini) のセクションにエントリが定義されていません。%1 セクションに %2 エントリを正しく定義する必要があります。

%1 : セクション名

%2 : エントリ名

(S)

処理を終了します。

(O)

%1 セクションに %2 エントリを定義してください。

KMBV12251-W <F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

6. トラブルシュート機能

エラー情報 : %2

ファイル名 : %3

An error occurred in extracting text data from document(%1).

Error information : %2

File name : %3

文書 (%1) のテキスト抽出処理中にテキスト抽出フィルタプラグインライブラリでエラーが発生したため、テキスト抽出に失敗しました。

%2 : エラー情報

%3 : ファイル名

エラー情報 (%2) でエラー内容を確認してください。

(S)

処理を続行します。

(O)

表 6-14 に示す対策一覧を基にエラーの要因を取り除いてください。

表 6-14 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
File write error. Reason code = xx	環境変数 TMPDIR 下に一時ファイル (%3) の作成に失敗しました。	Reason code (xx) を基にエラーの要因を取り除いてから、処理を再実行してください。 xx : システムコールのエラー番号 エラー番号については、OS のマニュアルを参照してください。
File read error. Reason code = xx	プラグインライブラリの関数が出力したテキスト抽出後のファイル (%3) の読み込みに失敗しました。	
File unlink error. Reason code = xx	環境変数 TMPDIR 下に作成した一時ファイル、またはプラグインライブラリの関数が出力したテキスト抽出後のファイル (%3) の削除処理に失敗しました。	

KMBV12252-W <F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

エラー情報 : %2

ライブラリ名 : %3

関数名 : %4

An error occurred in extracting text data from document(%1).

Error information : %2

Library name : %3

Function name : %4

文書 (%1) のテキスト抽出処理中にテキスト抽出フィルタプラグインライブラリでエラーが発生したため、テキスト抽出に失敗しました。

%2 : エラー情報

%3 : ライブラリ名

%4 : 関数名。エラー情報が「PlugIn library loading failed.」の場合は「-」を表示します。

エラー情報 (%2) でエラー内容を確認してください。

(S)

処理を続行します。

(O)

表 6-15 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-15 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
PlugIn library loading failed.	プラグインライブラリのローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
Function of PlugIn library loading failed.	フィルタプラグイン関数のローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
PlugIn error code = xx	フィルタプラグイン関数でエラーが発生しました。 xx : フィルタプラグイン関数の戻り値	フィルタプラグイン関数の戻り値 (xx) を基にエラーの要因を取り除いてください。

KMBV12253-I <F,U>

フィルタプラグイン関数 (dbrldxTextFilter) のリターン情報。

データモード : %1

テキストデータ長 : %2

ファイル名 : %3

削除モード : %4

Return information of Filter PlugIn function (dbrldxTextFilter)

Data mode : %1

Registration size : %2

File name : %3

Delete mode : %4

6. トラブルシュート機能

フィルタプラグイン関数 (dbrIdxTextFilter) のリターン情報を表示します。

- %1 : 「FILE」または「MEMORY」を表示します。
- %2 : データモードが「MEMORY」の場合、フィルタプラグイン関数 (dbrIdxTextFilter) の出力引数 TextLength の値を表示します。データモードが「FILE」の場合、フィルタプラグイン関数 (dbrIdxTextFilter) の出力引数 TextData に指定したファイルのサイズを表示します。
- %3 : データモードが「MEMORY」の場合、「-」を表示します。データモードが「FILE」の場合、フィルタプラグイン関数 (dbrIdxTextFilter) の出力引数 TextData に指定したファイル名を表示します。
- %4 : データモードが「MEMORY」の場合、「-」を表示します。データモードが「FILE」の場合、フィルタプラグイン関数 (dbrIdxTextFilter) の出力引数 OutFileDeleteFlag に指定した値を表示します。

(S)

処理を続行します。

(O)

-

KMBV12254-W <F,U>

文書 (%1) のプロパティ値の取得用処理が失敗しました。

エラー情報 : %2

ライブラリ名 : %3

関数名 : %4

An error occurred in processing for getting of the property value of document (%1) .

Error information : %2

Library name : %3

Function name : %4

文書 (%1) のプロパティ値取得用プラグインライブラリでエラーが発生したため、プロパティ値の取得処理に失敗しました。

%2 : エラー情報

%3 : ライブラリ名

%4 : 関数名。エラー情報が「PlugIn library loading failed.」の場合は「-」を表示します。

エラー情報 (%2) でエラー内容を確認してください。

(S)

処理を続行します。

(O)

表 6-16 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-16 要因別の対策一覧

エラー情報 (%2) に示さ れるメッセー ジ	エラーの要因	対策
PlugIn library loading failed.	プラグインライブラリのローディングに失 敗しました。	プラグイン環境設定ファイルの指定を 見直してください。
Function of PlugIn library loading failed.	プラグイン関数のローディングに失敗しま した。	プラグイン環境設定ファイルの指定を 見直してください。
PlugIn error code = xx	プラグイン関数でエラーが発生しました。 xx : プラグイン関数の戻り値	プラグイン関数の戻り値 (xx) を基に エラーの要因を取り除いてください。

KMBV12301-E <C,P,F,U>

構造指定全文検索ファイル作成時にエラーが発生しました。

エラー情報 : %1

An error occurred making struct all sentence search file.

Error information : %1

構造指定全文検索ファイル作成時に、エラー情報 (%1) に示すエラーが発生しました。

(S)

処理を終了します。

(O)

表 6-17 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-17 要因別の対策一覧

エラー情報 (%1) に示 されるメッセージ	エラーの要因	対策
HiRDB Adapter for XML loading failed.	次のどちらかの要因が考えられます。 <ul style="list-style-type: none"> • HiRDB Adapter for XML がインス トールされていない • 環境変数 PATH にパスが設定されて いない 	要因に応じて次のどちらかの対 策を実施してください。 <ul style="list-style-type: none"> • HiRDB Adapter for XML をイ ンストールする • 環境変数 PATH のパスを確認 する

KMBV12302-E <C,P,F,U>

構造指定全文検索ファイル作成時にエラーが発生しました。

エラー情報 : %1

6. トラブルシュート機能

An error occurred making struct all sentence search file.

Error information : %1

構造指定全文検索ファイル作成時に、エラー情報（%1）に示すエラーが発生しました。

（S）

処理を終了します。

（O）

表 6-18 に示す対策一覧を基にエラーの要因を取り除いてください。

表 6-18 要因別の対策一覧

エラー情報（%1）に示されるメッセージ	エラーの要因	対策
XMS file writing error.	環境変数 TMPDIR 下に XMS ファイル作成に失敗しました。	環境変数 TMPDIR のアクセス権限、ディスク容量を確認してください。
TmpDir access error.	環境変数 TMPDIR 下の一時ファイルの読み込みに失敗しました。	
TmpFile create error.	環境変数 TMPDIR 下に全文検索インデクス一時保管用ファイルの作成に失敗しました。	
Content File writing error.	環境変数 TMPDIR 下に文書実体一時保管用ファイルの作成に失敗しました。	
Preprocessing Library for Text Search loading failed.	Preprocessing Library for Text Search がインストールされていません。	Preprocessing Library for Text Search をインストールしてください。
xx	全文検索インデクスファイルの作成に失敗しました。	エラー情報（%1）に従って、エラーの要因を取り除いてください。

注 Preprocessing Library for Text から返されたエラーメッセージ

KMBV12303-E <C,P,F,U>

文書（%1）のテキストデータ登録時にエラーが発生しました。

要因コード：%2

テキストデータ長：%3

Error occurred in entry text data from document（%1）.

Reason code：%2

Length of text data：%3

文書（%1）のテキストデータ登録時にエラーが発生しました。テキストデータ長は%3バイトです。要因コード（%2）でエラーの内容を確認してください。

(S)

処理を終了します。

(O)

表 6-19 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-19 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	XML 文書のテキストデータのデータ長が最大値を超えたため、テキストデータは登録できませんでした。	MaxOfLength に設定した値をテキストデータ長以上にして、再実行してください。

KMBV12305-E <C,P,F,U>

文書 (%1) のテキストデータ抽出処理が失敗しました。

要因コード : %2

An error occurred in extracting text data from document (%1) .

Reason code : %2

文書 (%1) のテキスト抽出処理中にテキスト抽出ライブラリでエラーが発生したため、テキスト抽出処理が失敗しました。要因コード (%2) でエラーの内容を確認してください。

(S)

処理を終了します。

(O)

要因コード (%2) を基に、Document Filter for Text Search のドキュメントを参照し、対策を実施してください。

KMBV12306-E <C,P,F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

エラー情報 : %2

An error occurred in extracting text data from document (%1) .

Error information : %2

次のどちらかの要因で、テキスト抽出処理に失敗しました。

- 文書 (%1) のテキスト抽出処理中に、テキスト抽出ライブラリでエラーが発生した
- 環境変数 TMPDIR 下に、PDF 用一時ファイル作成に失敗した

エラー情報 (%2) でエラーの内容を確認してください。

(S)

6. トラブルシュート機能

処理を終了します。

(O)

表 6-20 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-20 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
PDF file writing error.	PDF の書き込みに失敗しました。	環境変数 TMPDIR のアクセス権限またはディスク容量を確認してください。
TmpDir access error.	一時ファイルの読み込みに失敗しました。	環境変数 TMPDIR のアクセス権限を確認してください。
Unmatch PDF file.	次のどちらかの要因によってテキスト抽出に失敗しました。 <ul style="list-style-type: none"> • 処理対象の PDF ファイルのバージョンがサポート対象外である • 処理対象の PDF ファイルの書式が不正 	PDF ファイルのバージョンおよび書式を確認してください。
Not PDF file.	PDF ファイル以外の文書が指定されています。	RenditionType 定義ファイルのフィルタ拡張子の定義が誤っていないか、または RenditionType の定義が誤っていないかどうかを確認してください。
PDF file Library error.	PDF テキスト抽出ライブラリでエラーが発生しました。	トレースログとエラーが発生した文書を採取して、保守員に連絡してください。

KMBV12307-E <C,P,F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

エラー情報 : %2

ファイル名 : %3

An error occurred in extracting text data from document (%1) .

Error information : %2

File name : %3

文書 (%1) のテキスト抽出処理中にテキスト抽出プラグインライブラリでエラーが発生したため、テキスト抽出処理に失敗しました。

エラー情報 (%2) でエラーの内容を確認してください。

%3 : ファイル名

(S)

処理を終了します。

(O)

表 6-21 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-21 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
File write error. Reason code = xx	環境変数 TMPDIR 下に一時ファイル (%3) の作成に失敗しました。	Reason code (xx) を基にエラーの要因を取り除いてから、処理を再実行してください。 xx : システムコールのエラー番号 エラー番号については、OS のマニュアルを参照してください。
File read error. Reason code = xx	プラグインライブラリの関数が出力したテキスト抽出後のファイル (%3) の読み込みに失敗しました。	
File unlink error. Reason code = xx	環境変数 TMPDIR 下に作成した一時ファイル、またはプラグインライブラリの関数が出力したテキスト抽出後のファイル (%3) の削除に失敗しました。	

KMBV12308-E <C,P,F,U>

文書 (%1) のテキスト抽出処理が失敗しました。

エラー情報 : %2

ライブラリ名 : %3

関数名 : %4

An error occurred in extracting text data from document (%1) .

Error information : %2

Library name : %3

Function name : %4

文書 (%1) のテキスト抽出処理中にテキスト抽出フィルタプラグインライブラリでエラーが発生したため、テキスト抽出処理に失敗しました。

エラー情報 (%2) でエラーの内容を確認してください。

%3 : ライブラリ名

%4 : 関数名。エラー情報が「PlugIn library loading failed.」の場合は「-」を表示しません。

エラー情報 (%2) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-22 に示す対策一覧を基に、エラーの要因を取り除いてください。

6. トラブルシュート機能

表 6-22 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
PlugIn library loading failed.	プラグインライブラリのローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
Function of PlugIn library loading failed.	プラグインライブラリの関数のローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
PlugIn error code = XX	プラグインライブラリの関数でエラーが発生しました。 XX : プラグインライブラリの関数の戻り値	プラグインライブラリの関数の戻り値 (XX) を基にエラーの要因を取り除いてください。
PlugIn error = Exception	プラグインライブラリの関数で例外が発生しました。	プラグインライブラリの関数では throw しないよう処理を見直してください。

KMBV12309-E <C,P,F,U>

文書 (%1) のプロパティ値の取得用処理が失敗しました。

エラー情報 : %2

ライブラリ名 : %3

関数名 : %4

An error occurred in processing for getting of the property value of document (%1) .

Error information : %2

Library name : %3

Function name : %4

文書 (%1) のプロパティ値の取得処理中に、プロパティ値取得用プラグインライブラリでエラーが発生したため、プロパティ値取得処理に失敗しました。

エラー情報 (%2) でエラーの内容を確認してください。

%3 : ライブラリ名

%4 : 関数名。エラー情報が「PlugIn library loading failed.」の場合は「-」を表示しません。

エラー情報 (%2) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-23 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-23 要因別の対策一覧

エラー情報 (%2) に示されるメッセージ	エラーの要因	対策
PlugIn library loading failed.	プラグインライブラリのローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
Function of PlugIn library loading failed.	プラグインライブラリの関数のローディングに失敗しました。	プラグイン環境設定ファイルの指定を見直してください。
PlugIn error code = XX	プラグインライブラリの関数でエラーが発生しました。 XX : プラグインライブラリの関数の戻り値	プラグインライブラリの関数の戻り値 (XX) を基にエラーの要因を取り除いてください。
PlugIn error = Exception	プラグインライブラリの関数で例外が発生しました。	プラグインライブラリの関数では throw しないよう処理を見直してください。
PlugIn property getting failed.	プロパティ値取得用プラグインライブラリを呼ぶための前処理 (プロパティの取得処理) に失敗しました。	このメッセージの前に出力されたメッセージ KMBV12053-E を基に、エラーの要因を取り除いてください。

KMBV12900-I <C,P,F>

%1 コマンドを実行中です。

It is executing %1.

(S)

%1 コマンド実行中です。

(O)

-

KMBV12901-I <C,P,F>

%1 コマンドの処理が終了しました。

It is success to execute %1.

(S)

%1 コマンドの処理を終了します。

(O)

-

KMBV12904-E <C,P,F>

定義ファイルの生成でエラーが発生しました。

6. トラブルシュート機能

要因コード : %1

An error occurred creating of Define file.

Reason code : %1

定義ファイルの生成でエラーが発生しました。要因コード (%1) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

要因コード (%1) の内容を確認してエラーの要因を取り除いてから再実行してください。

表 6-24 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	ファイルパスが存在しません。	ファイルパスを確認してください。
2	パスの長さが制限値 (255 バイト) を超えています。	定義ファイルパスを確認してください。
3	ディレクトリまたはファイルへのアクセス権がありません。	ディレクトリまたはファイルへのアクセス権を確認してください。

KMBV12905-E <C,P,F>

データベースへの接続でエラーが発生しました。

An error occurred in connecting database.

次のどれかの要因でデータベースに接続できません。

- データベースサーバが起動していない
- DocumentBroker サーバの接続情報の定義を使用してデータベースに接続する場合に、DocumentSpace 構成定義ファイル (docspace.ini) が存在しないか、またはアクセス権がない
- 定義内容が誤っている

(S)

処理を終了します。

(O)

データベースの起動または接続情報の定義を確認してください。

KMBV12906-E <C,P,F>

生成対象のテーブルがありません。

There is no table.

コマンドの引数に指定したユーザ ID に、DocumentBroker サーバが提供するテーブルがありませんでした。

(S)

処理を終了します。

(O)

指定したユーザ ID を見直してください。

KMBV12907-E <P>

他のプログラムが実行中のため、コマンドが実行できません。

Unable to execute this command because other program is executing.

該当の機能を複数実行しようとしています。

(S)

処理を終了します。

(O)

DocumentBroker サーバを停止後、または実行中の処理が終了してから再度実行してください。

KMBV12908-E <P>

環境変数の指定に誤りがあります。

環境変数名 : %1

Environment variable is invalid.

Environment variable name : %1

環境変数 (%1) が定義されていないか、指定値が誤っているため、コマンドが実行できません。定義ファイル生成ツールの実行には、環境変数 TEXTINDEXDIR を正しく定義する必要があります。

(S)

処理を終了します。

(O)

正しくインストールされているかを確認してください。

KMBV12909-E <P>

指定したコマンドの形式に誤りがあります。コマンドの形式は次のとおりです。

EDMCrtTIDF ユーザ ID [-d 定義ファイル]

Error exists in command line, usage is as follows

6. トラブルシュート機能

Usage: EDMCrtTIDF USER_ID [-d DefineFile]

指定したコマンド形式に誤りがあります。

(S)

処理を終了します。

(O)

-

KMBV12910-W <C,P,F>

メタ情報 (%1) に不正な定義があるため、ファイルに出力できませんでした。

Invalid definition in META (%1)

メタ情報 (%1) は、DisplayName と Guid 値の数が合っていないため処理できません。

(S)

エラーが発生したメタ情報を無視して処理を続行します。

(O)

メタ情報ファイルの定義内容、およびデータベースの定義を見直してください。

KMBV12990-E <C,P,F>

メモリが不足しているため、処理を続行できません。

There is insufficient memory.

メモリが不足しました。

(S)

処理を終了します。

(O)

不用なプログラムを終了させて再度実行してください。

KMBV12991-E <C,P,F>

ディスク容量が不足しているため、処理を続行できません。

There is insufficient space of disk.

ディスク容量が不足しました。

(S)

処理を終了します。

(O)

ディスクに空きがあることを確認してから再度実行してください。

KMBV12992-E <C,P,F>

%1 で内部矛盾が発生しました。

エラー情報 : %2

Internal error occurred to %1.

Error information : %2

 %1 で内部矛盾が発生しました。

 %2 : エラー情報

(S)

 処理を終了します。

(O)

 トレースログを採取して保守員に連絡してください。

KMBV12993-E <C,P,F>

データベースでエラーが発生しました。

エラー情報 : %1

要因コード : %2

An error occurred in database.

Error information : %1

Reason code : %2

 データベースへのアクセスでエラーが発生しました。

 %1 : データベースのエラーメッセージ

 %2 : データベースのエラーコード

(S)

 処理を終了します。

(O)

 エラーの要因を取り除いてください。

KMBV12994-E <C,P,F>

システムコールでエラーが発生しました。

関数名 : %1

要因コード : %2

An error occurred on the system call function.

Function Name : %1

6. トラブルシュート機能

Reason Code : %2

システムコール (%1) でエラーが発生しました。

%1 : エラーの発生した関数名

%2 : システムコールで返された errno 値

(S)

処理を終了します。

(O)

%2 に出力される要因コードを基に障害の要因を取り除いてください。

KMBV20000-W <C,P,F>

NTconfig.ini のエントリ %1 に指定したディレクトリが不正です。デフォルトディレクトリを使用します。

Invalid directory that specified %1 entry in NTconfig.ini. This Program uses default directory.

NTconfig.ini のエントリ %1 に指定したディレクトリが不正です。デフォルトディレクトリを使用します。

(S)

処理を続行します。

(O)

-

KMBV20001-W <C,P,F>

NTconfig.ini にエントリ %1 が複数存在します。最上位の指定を有効にします。

Same entry %1 exists in NTconfig.ini. First entry is used.

NTconfig.ini にエントリ %1 が複数存在します。最上位の指定を有効にします。

(S)

処理を続行します。

(O)

-

KMBV20002-E <C,P>

NTconfig.ini ファイル (%1) のオープンでエラーが発生しました。

要因コード : %2

An error occurred in opening of NTconfig.ini file(%1).

Reason code : %2

NTconfig.ini ファイル (%1) のオープンでエラーが発生しました。
 要因コード (%2) でエラー内容を確認してください。

(S)

処理を終了します。

(O)

表 6-25 に示す対策一覧を基に、エラーの要因を取り除いてください。

表 6-25 要因コード別の対策一覧

要因コード	エラーの要因	対策
1	アクセス権限がありません。	NTconfig.ini ファイルのアクセス権限を確認してください。
2	ファイルが存在しません。	次のどちらかの方法で対処してください。 <ul style="list-style-type: none"> システム環境変数に環境変数 TEXTINDEXDIR が設定されていることを確認してください。 <環境変数 TEXTINDEXDIR に指定されているインストールディレクトリ> ¥sample に存在する NTconfig.ini ファイルを <環境変数 TEXTINDEXDIR に指定されているインストールディレクトリ> ¥etc にコピーし NTconfig.ini ファイルを回復させてください。
3	パス名が 255 バイトを超えています。	NTconfig.ini ファイルのパス名が 255 バイト以下となるディレクトリにインストールし直してください。
4(%3)	システムコールでエラーが発生しました。 %3 : システムコールで返された errno	%3 を基に要因を取り除いてください。
5	メモリ不足が発生しました。	ほかに実行中のプログラムを終了させ、再度実行してください。
6	内部矛盾が発生しました。	保守員に連絡してください。

付録

付録 A 処理対象の全文検索インデクス用プロパティ

付録 B 文書空間で使用する文字コード種別が UTF-8 の場合のサポート範囲

付録 C 用語解説

付録 A 処理対象の全文検索インデクス用プロパティ

テキストインデクサーが処理対象とする全文検索インデクス用のプロパティ種別に対応する検索方法と文書種別を表 A-1 に示します。

表 A-1 テキストインデクサーが処理対象とする全文検索インデクス用プロパティ

運用	検索方法	プロパティ種別	対象	扱える文書の種別	内容
全文検索	ブレンテキスト検索	edmProp_Content ¹ edmProp_TextIndex		テキスト抽出ライブラリがサポートする文書種別 ²	ブレンテキストに対する概念検索を含まない全文検索
			×	XML 文書	-
	構造指定検索	edmProp_Content ¹ edmProp_StIndex		XML 文書	概念検索を含まない構造指定全文検索
概念検索	ブレンテキスト検索	edmProp_ConceptTextIndex		テキスト抽出ライブラリがサポートする文書種別 ²	ブレンテキストに対する概念検索を含まない全文検索
			×	XML 文書	-
	構造指定検索	edmProp_ConceptStIndex		XML 文書	概念検索を含まない構造指定全文検索

(凡例)

- : テキストインデクサーの対象にする
- ×: テキストインデクサーの対象にしない
- : 該当しない

注 1

edmProp_Content を使用している場合、該当する SGMLTEXT 型列定義のプラグインのテキストデータ登録種別が <TEXTTYPE>PLAIN</TEXTTYPE> のときだけ、テキストインデクサーの処理対象になります。

注 2

テキスト抽出ライブラリがサポートする文書種別については、使用しているテキスト抽出ライブラリの仕様を確認してください。

付録 B 文書空間で使用する文字コード種別が UTF-8 の場合のサポート範囲

ここでは、文書空間で使用する文字コード種別が UTF-8 の場合にサポートする範囲について説明します。なお、文書空間で使用する文字コード種別が Shift-JIS の場合は、すべての機能をサポートしています。

文書空間で使用する文字コード種別が UTF-8 の場合に使用できるテキスト抽出ライブラリを次の表に示します。

表 B-1 文書空間で使用する文字コード種別が UTF-8 の場合に使用できるテキスト抽出ライブラリ

テキスト抽出ライブラリ名	使用可否
Document Filter for Text Search	
HiRDB Adapter for XML	-
Hitachi Processing Facility for XML	-
フィルタプラグインライブラリ	-

(凡例)

- : 使用できます。
- : 使用できません。

注

Document Filter for Text Search のコンフィグレーションファイル (config.cfg) の設定で、出力するテキスト抽出結果のコード体系を「UTF-8」に設定して使用してください。

なお、テキスト抽出ライブラリを使用しないで、文書のテキスト情報を抽出することもできます。その場合、テキスト抽出結果の文字コード種別には UTF-8 を使用してください。

付録 C 用語解説

テキストインデクサーで使用する用語について説明します。

(英字)

DocumentSpace 構成定義ファイル (docspace.ini)

文書空間の構成を定義するためのファイルです。

(サ行)

書誌情報

文書コンテンツからのテキスト抽出失敗時に、全文検索用インデクスファイルを生成するために使用するプロパティを指します。

処理対象文書クラス

全文検索機能付き文書クラスおよび概念検索機能付き文書クラスを指します。

全文検索インデクス

全文検索の対象になるテキストデータに対応するプロパティです。

全文検索プラグイン

HiRDB Text Search Plug-in を指します。

(タ行)

テキスト抽出フィルタプラグインライブラリ

テキストインデクサーが定義する関数に従ってユーザが作成した、テキスト抽出ライブラリのことです。

テキスト抽出ライブラリ

Document Filter for Text Search ,Preprocessing Library for Text Search ,および HiRDB Adapter for XML を指します。各種アプリケーションデータから、全文検索用のテキストを自動的に抽出するプログラムプロダクトです。

(ラ行)

レンディション

文書のコンテンツの形式およびそのコンテンツをあわせてレンディションと呼びます。

レンディションタイプ

Word などのアプリケーションで編集したファイル、HTML 形式のファイルや GIF などの画像デー

タファイルのように、登録した文書のコンテンツ（データ部分）のファイル形式を表す文字列です。

索引

記号

_HIEDMS_TRACE_DIR 49
_HIEDMS_TRACE_LEVEL 50
_HIEDMS_TRACE_NUM 49
_HIEDMS_TRACE_SIZE 49

C

ClassNameDefinition 44

D

dbrIdxTextFilter() 73
DbType 37
Definition_def 33
DOCBROKERDIR 18
DocSpace_def 36
DocSpaceId 37
DocumentBroker サーバの設定 18
Document Filter for Text Search のパス設定
について 23
DocumentSpace 構成定義ファイル 136

E

EDM_Idx_com.h 68
EDMCrtTIDF 65
EDMTextIndex 61
EDMTextIndex コマンド実行時に -o オプ
ションを指定した場合 27
ErrorLog 35

H

HiRDB Text Search Plug-in での環境設定
20
HiRDB の環境の設定 20
HiRDB の起動 21

L

LimitTime 34

M

MaxOfLength 34
MaxOfLength とテキスト抽出の結果につい
て 22
ModeOfIndexing 38
ModeOfProperty 35

N

NTconfig.ini 16

P

PDF を示す RenditionType の値に [filterA]
または [filterC] のフィルタ拡張子を付けて実
行したとき 22
PdHost 37
PdNamePort 37
PDSWAITTIME 20
PdTSPluginOwner 38
PdUser 38
PlugInconfig.ini 16
PropertyOfIndex 35
PropNameDefinition 45

R

RenditionType.ini 16
RenditionType_def 34
RenditionType 定義ファイル 7, 41
RenditionType 定義ファイルの記述例 43

S

SystemObjectID 36

T

TextIndex.ini 16
TMPDIR 50
TMPDIR パラメタ 22

X

XmlBroker 36

あ

アンインストール 16

い

インストール 13

インストールディレクトリと提供ファイルの内容 14

インストールとアンインストール 13

え

エラー情報およびトレース情報が取得されるファイル 8

エラーログファイル 8, 80

エントリ 31

か

確認する環境変数 18

環境設定ファイル 8, 33

環境設定ファイルの記述例 39

環境定義パラメタ 33

環境変数の設定 18

環境変数の設定場所 18

環境変数ファイル 8, 48

環境変数ファイルに設定するパラメタ 49

く

クライアント側の環境設定 20

こ

効率的なインデクス登録処理について 21

コマンドの形式 59

コンパイルオプション 68

さ

サブレンディションの文書からのテキスト抽出について 25

し

システム構成 3

実行環境および運用に関する注意事項 21

実行環境設定の流れ 12

実行環境の設定 11

実行コマンド一覧 58

障害処理 81

使用するファイルの一覧 30

書誌情報 136

処理対象の全文検索インデクス用プロパティ 134

処理対象文書クラス 136

処理対象文書クラス定義ファイル 8, 46

処理対象文書クラス定義ファイルの記述例 46

処理文書長の上限値 81

せ

セクション 31

説明する項目 72

全体の処理時間の上限値 81

前提 PP 構成一覧 4

全文検索インデクス 136

全文検索インデクスを登録する場合の認可識別子 18

全文検索プラグイン 136

全文検索用インデクスファイル 8

全文検索用の表の作成 20

ち

抽出処理が失敗した文書の取り扱い 39

て

定義ファイル 8, 44

定義ファイル生成コマンドを使用して定義ファイルを生成した場合の注意事項 23

定義ファイルの記述例 45

提供するフィルタ拡張子と使用するテキスト抽出ライブラリ 41

提供ファイル 16

テキストインデクサー (EDMTextIndex) の
 テキスト抽出処理 21
 テキストインデクサー (EDMTextIndex) の
 並列実行 21
 テキストインデクサー, DocumentBroker お
 よび HiRDB の実行環境 23
 テキストインデクサーが LimitTime タイム
 アウトで終了したとき 22
 テキストインデクサーが対象とする文書クラ
 ス, および全文検索インデクスとして指定で
 けるプロパティについて 23
 テキストインデクサーが定義するプラグイン
 関数 68
 テキストインデクサーが提供するコマンド 5
 テキストインデクサーが提供するコマンドの
 一覧 5
 テキストインデクサーで使用するファイル
 29
 テキストインデクサーで使用できる文字コー
 ド種別 9
 テキストインデクサーとは 2
 テキストインデクサーの概要 1
 テキストインデクサーの環境変数の設定 19
 テキストインデクサーの機能 5
 テキストインデクサーのコマンド実行時間
 22
 テキストインデクサーの実行環境を作成する
 ディレクトリについて 21
 テキストインデクサーの実行者 21
 テキストインデクサーの処理概要 6
 テキストインデクサーの動作環境に関する注
 意 4
 テキストインデクサーを実行するために必要
 なファイル 7
 テキストインデクサーを実行する場合のテキ
 スト抽出用の一時ファイル作成ディレクトリ
 について 22
 テキスト情報抽出処理中にエラーが発生した
 場合 81
 テキスト抽出時の注意事項 24
 テキスト抽出フィルタプラグインライブラリ
 136
 テキスト抽出ライブラリ 136

と

トラブルシュート 79
 トラブルシュート機能の概要 80
 トレースログファイル 8, 80

に

入力文書ファイル 8

は

パスの指定方法 31

ふ

ファイル環境 7
 ファイル形式 31
 ファイル形式と文法 31
 フィルタプラグイン関数 73
 フィルタプラグイン機能 5
 プラグイン環境設定ファイル 8, 52
 プラグイン環境設定ファイルの記述規則 52
 プラグイン環境設定ファイルの記述例 55
 プラグイン関数一覧 68
 プラグイン関数の処理 70
 プラグイン関数の呼び出し規約 68
 プロパティ取得関数 76
 文書空間で使用する文字コード種別が UTF-
 8 の場合のサポート範囲 135
 文法 31

へ

ヘッダファイル 68

め

メッセージ ID の記号の説明 82
 メッセージの記述形式 82
 メッセージの形式 82
 メッセージの出力形式 82
 メッセージの詳細 84

れ

レンディション 136

レンディションタイプ 136

ソフトウェアマニュアルのサービス ご案内

1. マニュアル情報ホームページ

ソフトウェアマニュアルの情報をインターネットで公開しています。

URL <http://www.hitachi.co.jp/soft/manual/>

ホームページのメニューは次のとおりです。

マニュアル一覧	日立コンピュータ製品マニュアルを製品カテゴリ、マニュアル名称、資料番号のいずれかから検索できます。
CD-ROMマニュアル	日立ソフトウェアマニュアルと製品群別CD-ROMマニュアルの仕様について記載しています。
マニュアルのご購入	マニュアルご購入時のお申し込み方法を記載しています。
オンラインマニュアル	一部製品のマニュアルをインターネットで公開しています。
サポートサービス	ソフトウェアサポートサービスお客様向けページでのマニュアル公開サービスを記載しています。
ご意見・お問い合わせ	マニュアルに関するご意見、ご要望をお寄せください。

2. インターネットでのマニュアル公開

2種類のマニュアル公開サービスを実施しています。

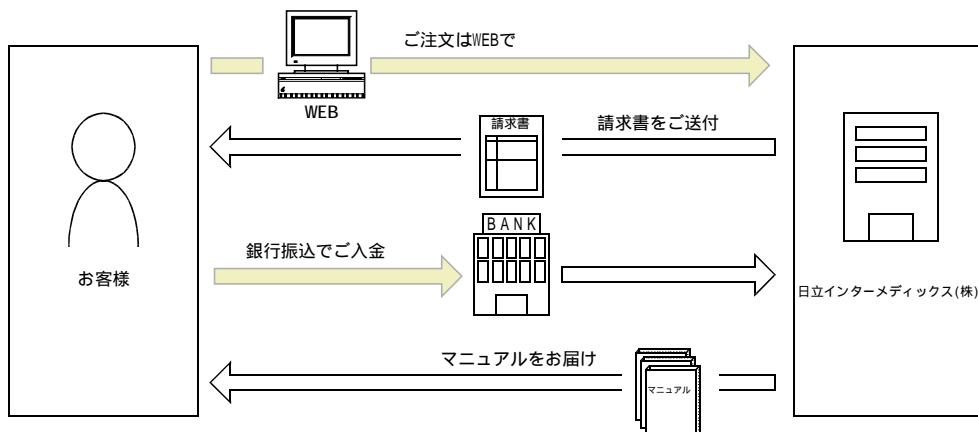
(1) マニュアル情報ホームページ「オンラインマニュアル」での公開

製品をよりご理解いただくためのご参考として、一部製品のマニュアルを公開しています。

(2) ソフトウェアサポートサービスお客様向けページでのマニュアル公開

ソフトウェアサポートサービスご契約のお客様向けにマニュアルを公開しています。公開しているマニュアルの一覧、本サービスの対象となる契約の種別などはマニュアル情報ホームページの「サポートサービス」をご参照ください。

3. マニュアルのご注文



マニュアル情報ホームページの「マニュアルのご購入」にアクセスし、お申し込み方法をご確認のうえ WEB からご注文ください。ご注文先は日立インターメディアックス(株)となります。

ご注文いただいたマニュアルについて請求書をお送りします。

請求書の金額を指定銀行へ振り込んでください。

入金確認後 7 日以内にお届けします。在庫切れの場合は、納期を別途ご案内いたします。