HITACHI
Inspire the Next

Document Filter for Text Search Version 3

解説・手引書

3020-3-D59-60

■対象製品

P-24D1-2F34 Document Filter for Text Search Version 3 03-42（適用 OS：Windows 2000，Windows Server 2003，Windows Server 2003 x64 Edition，Windows Server 2003 R2，Windows Server 2003 R2 x64 Edition，Windows Server 2008，Windows Server 2008 x64 Edition，Windows Server 2008 R2）
P-1MD1-2F31 Document Filter for Text Search Version 3 03-42（適用 OS：AIX 5L V5.2，AIX 5L V5.3，AIX V6.1）※1※3
P-1BD1-2F31 Document Filter for Text Search Version 3 03-42（適用 OS：HP-UX 11，11i）※1※2
P-9DD1-2F31 Document Filter for Text Search Version 3 03-42（適用 OS：Solaris 8，Solaris 9，Solaris 10）※4
※1 これらの製品については，サポート時期をご確認ください。
※2 HP-UX 11 で使用する場合は，パッチの適用が必要なときがあります。詳細は「リリースノート」をご覧ください。
※3 AIX 5L V5.3 の場合，パッチの適用が必要なときがあります。詳細は「リリースノート」をご覧ください。
※4 Solaris 8，Solaris 9 の場合，パッチの適用が必要なときがあります。詳細は「リリースノート」をご覧ください。
上記のプログラムプロダクトのほかにもこのマニュアルをご利用になれる場合があります。詳細は「リリースノート」でご確認ください。

■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

■マイクロソフト製品の表記について

このマニュアルでは，マイクロソフト製品の名称を次のように表記しています。

表記	製品名
Excel	Microsoft(R) Excel，Microsoft(R) Office Excel(R)
Power Point	Microsoft(R) PowerPoint，Microsoft(R) Office PowerPoint(R)
Word	Microsoft(R) Word，Microsoft(R) Office Word

表記		製品名
Windows	Windows 2000	Microsoft(R) Windows(R) 2000 Advanced Server Operating System
		Microsoft(R) Windows(R) 2000 Datacenter Server Operating System
		Microsoft(R) Windows(R) 2000 Professional Operating System
	Windows Server 2003	Microsoft(R) Windows Server(R) 2003, Enterprise Edition
		Microsoft(R) Windows Server(R) 2003, Standard Edition
	Windows Server 2003 R2	Microsoft(R) Windows Server(R) 2003 R2, Enterprise Edition
		Microsoft(R) Windows Server(R) 2003 R2, Standard Edition
	Windows Server 2003 x64	Microsoft(R) Windows Server(R) Enterprise x64 Edition
		Microsoft(R) Windows Server(R) Standard x64 Edition
		Microsoft(R) Windows Server(R) R2, Enterprise x64 Edition
	Windows Server 2003 x86	Microsoft(R) Windows Server(R) 2003, Enterprise x86 Edition
		Microsoft(R) Windows Server(R) 2003, Standard x86 Edition
		Microsoft(R) Windows Server(R) 2003 R2, Enterprise x86 Edition
		Microsoft(R) Windows Server(R) 2003 R2, Standard x86 Edition
	Windows Server 2008	Microsoft(R) Windows Server(R) 2008, Enterprise Edition
		Microsoft(R) Windows Server(R) 2008, Standard Edition
		Microsoft(R) Windows Server(R) 2008 Datacenter
	Windows Server 2008 x64	Microsoft(R) Windows Server(R) 2008, Enterprise x64 Edition
		Microsoft(R) Windows Server(R) 2008, Standard x64 Edition
		Microsoft(R) Windows Server(R) 2008 Datacenter x64 Edition
	Windows Server 2008 R2	Microsoft(R) Windows Server(R) 2008 R2, Enterprise Edition
		Microsoft(R) Windows Server(R) 2008 R2, Standard Edition

■発行

2012 年 5 月（第 8 版）3020-3-D59-60

変更内容

変更内容（3020-3-D59-60） Document Filter for Text Search Version 3　03-42

追加・変更内容	変更個所
使用できる上位アプリケーションに次の製品を追加しました。 • DocumentBroker Text Search Index Loader Version 3 • Enterprise Search を追加しました。	1.1，B.1
使用するフィルタに MS64 を追加しました。	1.3，1.4，2.3.1，2.4，2.4.2

単なる誤字・脱字などはお断りなく訂正しました。
なお，「はじめに」の記載の一部を「このマニュアルの参考情報」に移動しました。

はじめに

このマニュアルは，プログラムプロダクト Document Filter for Text Search Version 3（Document Filter for Text Search）の機能，環境設定方法および使用方法について説明したものです。

■対象読者

このマニュアルは，Document Filter for Text Search の環境を，管理および運用するシステム管理者を対象に説明しています。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。

- 全文検索用のテキスト抽出に関する基本的な知識
- テキスト抽出の対象となるドキュメントに関する基本的な知識
- Document Filter for Text Search を使用する上位アプリケーションに関する基本的な知識
- Windows，または UNIX（AIX，HP-UX，Solaris）に関する基本的な知識

■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項目は省略できることを示します。 （例）［A］ 「何も指定しない」か「A を指定する」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。

目次

1 概要 1

1.1 Document Filter for Text Search とは 2

1.1.1 システム構成 2

1.2 処理概要 3

1.3 対象ドキュメント 4

1.4 抽出するプロパティ情報 5

2 環境設定 7

2.1 環境設定の流れ 8

2.2 インストールとアンインストール 9

2.2.1 インストールの方法 9

2.2.2 アンインストールの方法 12

2.3 コンフィグレーションファイルの設定 15

2.3.1 定義内容 15

2.3.2 記述規則 18

2.3.3 定義例 18

2.4 テキスト抽出時の注意事項 19

2.4.1 DMC フィルタの場合 22

2.4.2 MSIF/MS64 フィルタの場合 24

2.4.3 DOCF フィルタの場合 25

2.4.4 DMTX フィルタの場合 25

3 障害対策 27

3.1 障害情報の取得 28

3.2 詳細情報ファイル 29

3.3 一時ファイル 30

3.4 詳細コード 31

付録 33

付録 A フォルダ構成 34

付録 B このマニュアルの参考情報 35

付録 B.1　関連マニュアル　35
付録 B.2　このマニュアルでの表記　35
付録 B.3　英略語　36
付録 B.4　KB（キロバイト）などの単位表記について　36
付録 C　用語解説　37

索引　39

1 概要

この章では，Document Filter for Text Search の機能，処理概要，および対象ドキュメントについて説明します。

1.1 Document Filter for Text Search とは

1.2 処理概要

1.3 対象ドキュメント

1.4 抽出するプロパティ情報

1.1 Document Filter for Text Search とは

現在のオフィスには，ワープロやスプレッドシートなど，さまざまなアプリケーションで作成された文書が散在しています。Document Filter for Text Search は，各種アプリケーションで作成された文書から，全文検索用のテキストデータを抽出するためのユティリティプログラムです。この機能を**テキスト抽出機能**と呼びます。

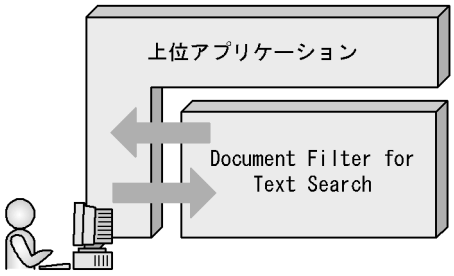
Document Filter for Text Search は，上位アプリケーションから実行します。使用できる上位アプリケーションは次の通りです。

- DocumentBroker Text Search Index Loader
- Groupmax Document Manager
- HiRDB Text Search Plug-in Index Generator
- Enterprise Search

1.1.1 システム構成

Document Filter for Text Search のシステム構成を次に示します。Document Filter for Text Search は，上位アプリケーションと同一マシン上で動作することが前提となります。

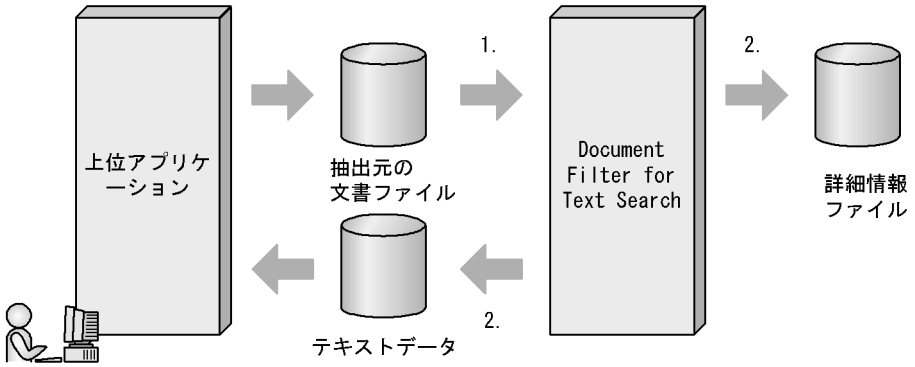
図 1-1 Document Filter for Text Search のシステム構成



1.2 処理概要

Document Filter for Text Search のテキスト抽出機能の処理概要を次に示します。

図 1-2 テキスト抽出機能の処理概要



1. Document Filter for Text Search は，上位アプリケーションの要求を受けて，テキストデータを抽出します。
2. Document Filter for Text Search は，抽出したテキストデータを上位アプリケーションに返します。また，詳細情報ファイルを出力します。

1.3 対象ドキュメント

Document Filter for Text Search がテキスト抽出の対象とするドキュメント（文書種別）を次に示します。各ドキュメントから抽出できるプロパティ情報については，「1.4 抽出するプロパティ情報」を参照してください。

表 1-1 対象ドキュメント

使用するフィルタ※1	文書種別
DMC	Word
	Excel
	PowerPoint
	一太郎
	PDF
	Lotus 1-2-3
	OASYS
	DocuWorks
	RTF
	HTML
	XML
MSIF※2※3	－（導入する IFilter に依存します）
MS64※2※4	－（導入する IFilter に依存します）
DOCF	テキスト（UCS-2 の範囲）
DMTX	テキスト（UCS-4 の範囲）

注※ 1

使用するフィルタは，コンフィグレーションファイルに設定します。コンフィグレーションファイルの詳細については，「2.3 コンフィグレーションファイルの設定」を参照してください。

注※ 2

Windows の Index Service のインタフェースに準拠したフィルタです。

注※ 3

64bit OS 環境では WOW64 で動作する IFilter を対象とします。

注※ 4

64bit OS 環境でだけ使用できます。

1.4 抽出するプロパティ情報

Document Filter for Text Search では，文書のプロパティ情報とテキストデータ（本文）を抽出します。抽出できるプロパティ情報は文書種別によって異なります。

抽出できるプロパティ情報とテキストデータを次に示します。

表 1-2 DMC フィルタで抽出できるプロパティ情報およびテキストデータ

文書種別	プロパティ情報						テキストデータ
	タイトル	サブタイトル	作成者	キーワード	リンク情報	見出し	
Word	○	○	○	○	×	×	○
Excel	○	○	○	○	×	×	○
PowerPoint	○	○	○	○	×	×	○
一太郎	○	×	○	○	×	×	○
PDF	○	○	○	○	×	×	○
Lotus 1-2-3	○	○	○	○	×	×	○
OASYS	○	○	○	○	×	×	○
DocuWorks	○	○	○	○	×	×	○
RTF	○	○	○	○	×	×	○
HTML	×	×	×	×	×	×	○
XML	×	×	×	×	×	×	○

（凡例）

○：文書実体ファイルから抽出できるプロパティ情報

×：文書実体ファイルから抽出できないプロパティ情報

表 1-3 MSIF/MS64 フィルタで抽出できるプロパティ情報およびテキストデータ

文書種別	プロパティ情報						テキストデータ
	タイトル	サブタイトル	作成者	キーワード	リンク情報	見出し	
任意	○	○	○	○	○	○	○

注　対応するプロパティの情報を取得しますが，実際に抽出できるかどうかは，インストールされた IFilter の仕様に依存します。

（凡例）

○：文書実体ファイルから抽出できるプロパティ情報

×：文書実体ファイルから抽出できないプロパティ情報

表 1-4　DOCF フィルタで抽出できるプロパティ情報およびテキストデータ

文書種別	プロパティ情報						テキストデータ
	タイトル	サブタイトル	作成者	キーワード	リンク情報	見出し	
テキスト	×	×	×	×	×	×	○

（凡例）
○：文書実体ファイルから抽出できるプロパティ情報
×：文書実体ファイルから抽出できないプロパティ情報

表 1-5　5DMTX フィルタで抽出できるプロパティ情報およびテキストデータ

文書種別	プロパティ情報						テキストデータ
	タイトル	サブタイトル	作成者	キーワード	リンク情報	見出し	
テキスト	×	×	×	×	×	×	○

（凡例）
○：文書実体ファイルから抽出できるプロパティ情報
×：文書実体ファイルから抽出できないプロパティ情報

2 環境設定

この章では，Document Filter For Text Search の環境設定の方法，およびテキスト抽出時の注意事項について説明します。

2.1 **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 **コンフィグレーションファイルの設定**

2.4 **テキスト抽出時の注意事項**

2.1 環境設定の流れ

Document Filter for Text Search の環境設定の流れを次に示します。

1. Document Filter for Text Search をインストールします。
   インストールとアンインストールの方法については,「2.2 インストールとアンインストール」を参照してください。
2. Document Filter for Text Search のテキスト抽出環境を設定します。
   テキスト抽出環境の設定方法については,「2.3 コンフィグレーションファイルの設定」を参照してください。

2.2 インストールとアンインストール

この節では，Document Filter for Text Search のインストールとアンインストールの方法について説明します。

2.2.1 インストールの方法

Document Filter for Text Search のインストールの手順を次に示します。

(1) [Windows]

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ます。
3. インストール用フロッピーディスクまたは CD-ROM 中の Setup.exe を起動します。
4. インストールプログラムが起動しますので，画面の指示に従ってインストールしてください。

(2) [AIX]

1. スーパーユーザ (id:root) としてオペレーティングシステムにログインします。
2. 「mkdir/cdrom」を実行して CD-ROM をマウントする「/cdrom」ディレクトリを作成します。
   「/cdrom」の部分には，CD-ROM をファイルシステムとしてマウントするディレクトリ名を指定してください。
3. Document Filter for Text Search の CD-ROM を CD-ROM ドライブにセットします。
4. 「mount -r -v cdrfs /dev/cd0 /cdrom」を実行して CD-ROM をマウントします。
   「/dev/cd0」および「/cdrom」の部分は使用する環境によって異なります。使用するデバイススペシャルファイル名および CD-ROM ファイルシステムを指定してください。
5. CD-ROM セットアッププログラム「/cdrom/aix/setup」を実行して日立 PP インストーラを起動します。
   初期画面が表示されます。なお，CD-ROM のディレクトリ名やファイル名は，ls コマンドで確認して表示されたファイル名を入力してください。
6. 「i」または「I」を入力して「I）Install Software」を選択します。
   インストールできるプログラム一覧が表示されます。
7. インストールするプログラムにカーソルを移動させて，スペースキーで選択します。
   選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
8. 「i」または「I」を入力して「I）Install」を選択します。
   プログラムをインストールするかどうかについてのメッセージが最下行に表示されま

す。

9. **最下行に表示されるメッセージに対して「y」または「Y」を入力します。**
   インストールが始まります。ただし，「n」または「N」を入力すると，インストールが中止されて，手順 6. で表示されたインストールできるプログラムの一覧が表示されます。

10. **プログラムのインストールが終了したら，「q」または「Q」を入力して「Q）Quit」を選択し，初期画面に戻ります。**
    インストールを終了する場合，初期画面で「q」または「Q」を入力して「Q）Quit」を選択して終了します。

（3）[HP-UX]

1. **スーパーユーザ（id:root）としてオペレーティングシステムにログインします。**
2. **「mkdir/cdrom」を実行して** CD-ROM **をマウントする「/cdrom」ディレクトリを作成します。**
   「/cdrom」の部分には，CD-ROM をファイルシステムとしてマウントするディレクトリ名を指定してください。
3. Document Filter for Text Search **の** CD-ROM **を** CD-ROM **ドライブにセットします。**
4. **「**mount -r -F cdfs /dev/dsk/c0t2d0 /cdrom**」を実行して** CD-ROM **をマウントします。**
   「/dev/dsk/c0t2d0」および「/cdrom」の部分は使用する環境によって異なります。使用するデバイススペシャルファイル名および CD-ROM ファイルシステムを指定してください。
5. CD-ROM **セットアッププログラム「**/cdrom/HPUX/SETUP**」を実行して日立** PP **インストーラを起動します。**
   初期画面が表示されます。なお，CD-ROM のディレクトリ名やファイル名は，ls コマンドで確認して表示されたファイル名を入力してください。
6. **「**i**」または「**I**」を入力して「**I）Install Software**」を選択します。**
   インストールできるプログラム一覧が表示されます。
7. **インストールするプログラムにカーソルを移動させて，スペースキーで選択します。**
   選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
8. **「**i**」または「**I**」を入力して「**I）Install**」を選択します。**
   プログラムをインストールするかどうかについてのメッセージが最下行に表示されます。
9. **最下行に表示されるメッセージに対して「y」または「Y」を入力します。**
   インストールが始まります。ただし，「n」または「N」を入力すると，インストールが中止されて，手順 6. で表示されたインストールできるプログラムの一覧が表示されます。

10. **プログラムのインストールが終了したら，「q」または「Q」を入力して「Q）Quit」を選択し，初期画面に戻ります。**
インストールを終了する場合，初期画面で「q」または「Q」を入力して「Q）Quit」を選択して終了します。

（4）［Solaris］

1. **スーパーユーザ（id:root）としてオペレーティングシステムにログインします。**
2. **Document Filter for Text Search の CD-ROM を CD-ROM ドライブにセットします。**
通常は自動マウントされますが，マウントされない場合は，手順 3. および手順 4. で CD-ROM をマウントします。
3. **「mkdir/cdrom」を実行して CD-ROM をマウントする「/cdrom」ディレクトリを作成します。**
「/cdrom」の部分には，CD-ROM をファイルシステムとしてマウントするディレクトリ名を指定してください。
4. **「mount -r -F hsfs /dev/dsk/c0t4d0s0 /cdrom」を実行して CD-ROM をマウントします。**
「/dev/dsk/c0t4d0s0」，および「/cdrom」の部分は使用する環境によって異なります。使用するデバイススペシャルファイル名，および CD-ROM ファイルシステムを指定してください。
5. **CD-ROM セットアッププログラム「/cdrom/Solaris/setup」を実行して日立 PP インストーラを起動します。**
初期画面が表示されます。なお，CD-ROM のディレクトリ名やファイル名は，ls コマンドで確認して表示されたファイル名を入力してください。
6. **「i」または「I」を入力して「I）Install Software」を選択します。**
インストールできるプログラム一覧が表示されます。
7. **インストールするプログラムにカーソルを移動させて，スペースキーで選択します。**
選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
8. **「i」または「I」を入力して「I）Install」を選択します。**
プログラムをインストールするかどうかについてのメッセージが最下行に表示されます。
9. **最下行に表示されるメッセージに対して「y」または「Y」を入力します。**
インストールが始まります。ただし，「n」または「N」を入力すると，インストールが中止されて，手順 6. で表示されたインストールできるプログラムの一覧が表示されます。
10. **プログラムのインストールが終了したら，「q」または「Q」を入力して「Q）Quit」を選択し，初期画面に戻ります。**
11. **インストールを終了する場合，初期画面で「q」または「Q」を入力して「Q）Quit」**

を選択して終了します。

2.2.2 アンインストールの方法

Document Filter for Text Search のアンインストールの手順を次に示します。

（1）［Windows］

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［アプリケーションの追加と削除］から，アンインストールを実行します。**

（2）［AIX］

1. **スーパーユーザ（**id:root**）としてオペレーティングシステムにログインします。**
2. **起動しているプログラムを停止します。**
3. **「**/etc/hitachi_setup**」を実行して日立** PP **インストーラを起動します。**
4. **「**d**」または「**D**」を入力して「**D**）**Delete Software**」を選択します。**
   アンインストールできるプログラムの一覧が表示されます。
5. **アンインストールするプログラムにカーソルを移動させて，スペースキーで選択します。**
   選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
6. **「**d**」または「**D**」を入力して「**D**）**Delete**」を選択します。**
   プログラムをアンインストールするかどうかについてのメッセージが最下行に表示されます。
7. **最下行に表示されるメッセージに対して「**y**」または「**Y**」を入力します。**
   アンインストールが始まります。ただし，「n」または「N」を入力すると，アンインストールが中止されて，手順 4. で表示されたアンイントールできるプログラム一覧が表示されます。
8. **プログラムのアンインストールが終了したら，「**q**」または「**Q**」を入力して「**Q**）Quit」を選択して初期画面に戻ります。**
   アンインストールを終了する場合，初期画面で「q」または「Q」を入力して「Q）Quit」を選択して終了します。

（3）［HP-UX］

1. **スーパーユーザ（**id:root**）としてオペレーティングシステムにログインします。**
2. **起動しているプログラムを停止します。**
3. **「**/etc/hitachi_setup**」を実行して日立** PP **インストーラを起動します。**

4. **「d」または「D」を入力して「D）Delete Software」を選択します。**
   アンインストールできるプログラムの一覧が表示されます。
5. **アンインストールするプログラムにカーソルを移動させて，スペースキーで選択します。**
   選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
6. **「d」または「D」を入力して「D）Delete」を選択します。**
   プログラムをアンインストールするかどうかについてのメッセージが最下行に表示されます。
7. **最下行に表示されるメッセージに対して「y」または「Y」を入力します。**
   アンインストールが始まります。ただし，「n」または「N」を入力すると，アンインストールが中止されて，手順 4. で表示されたアンイントールできるプログラム一覧が表示されます。
8. **プログラムのアンインストールが終了したら，「q」または「Q」を入力して「Q）Quit」を選択して初期画面に戻ります。**
   アンインストールを終了する場合，初期画面で「q」または「Q」を入力して「Q）Quit」を選択して終了します。

（4）［Solaris］

1. **スーパーユーザ（id:root）としてオペレーティングシステムにログインします。**
2. **起動しているプログラムを停止します。**
3. **「/etc/hitachi_setup」を実行して日立 PP インストーラを起動します。**
4. **「d」または「D」を入力して「D）Delete Software」を選択します。**
   アンインストールできるプログラムの一覧が表示されます。
5. **アンインストールするプログラムにカーソルを移動させて，スペースキーで選択します。**
   選択したプログラムの左側に「<@>」が付きます。なお，複数のプログラムを選択できます。
6. **「d」または「D」を入力して「D）Delete」を選択します。**
   プログラムをアンインストールするかどうかについてのメッセージが最下行に表示されます。
7. **最下行に表示されるメッセージに対して「y」または「Y」を入力します。**
   アンインストールが始まります。ただし，「n」または「N」を入力すると，アンインストールが中止されて，手順 4. で表示されたアンイントールできるプログラム一覧が表示されます。
8. **プログラムのアンインストールが終了したら，「q」または「Q」を入力して「Q）Quit」を選択して初期画面に戻ります。**

アンインストールを終了する場合，初期画面で「q」または「Q」を入力して「Q) Quit」を選択して終了します。

2.3 コンフィグレーションファイルの設定

Document Filter for Text Search のテキスト抽出環境を，コンフィグレーションファイル（config.cfg）に設定します。上位アプリケーションにテキスト抽出環境が設定されている場合は，上位アプリケーションの設定が優先されます。

この節では，コンフィグレーションファイルの定義内容，記述規則，および定義例について説明します。

2.3.1 定義内容

コンフィグレーションファイルの定義内容を次に示します。

表 2-1 コンフィグレーションファイルの定義内容

セクション	エントリ	内容	設定値	デフォルト値
[SYSTEM]	DOCFINPUTCODE	テキスト抽出元のコード体系を示す識別子の文字列を指定します。 DOCF フィルタを使用して，テキスト形式ファイルからテキストを抽出する場合に有効になります。	• SJIS • JIS • EUC • UTF8	自動判定
	DOCFOUTPUTCODE	出力するテキスト抽出結果のコード体系を示す識別子の文字列を指定します。	• SJIS • JIS • EUC • UTF8	SJIS
	DOCFTIMEOUT	テキスト抽出時の 1 ファイルあたりのタイムアウト時間を指定します。	• 0 ～ 86400（秒） 0 の場合は，無制限になります。	0※1
	DOCFDELETEMODE	詳細情報の保管方法を指定します。 詳細情報を残すかどうか，および残すファイルを指定します。	• 0：保管しません。 • 1：原因不明のエラーで API がテキスト抽出に失敗した場合に，詳細情報ファイルを残します。 • 2：原因不明のエラーで API がテキスト抽出に失敗した場合に，詳細情報ファイルと一時ファイルを残します。	2

セクション	エントリ	内容	設定値	デフォルト値
	DOCFDAYTOLEAVE	DOCFDELETEMODE が 1 または 2 の場合に，保管した詳細情報の保存期間を日数で指定します。Document Filter for Text Search 実行時に，本エントリで指定した日数よりも古い詳細情報が存在する場合は削除します。	• 0 ～ 90（日） 0 の場合は無制限になります。	10
	DOCFWORKDIR	一時フォルダを設定します。 このフォルダには抽出時に一時ファイルが格納され，詳細情報の保管指定に従って保管されます。	一時ファイルの出力先をフルパスで指定します。 指定できるパス名は，最大 149 バイトまでです。	＜Document Filter for Text Search のインストールフォルダ＞¥tmp ※ 4
	DOCFATTACHED	添付ファイル付き文書の扱いを指定します。	• 0：未サポート文書と判断し，エラーとします。 • 1：添付ファイルを含めてテキスト抽出し，正常終了します。 • 2：元文書だけをテキスト抽出し，正常終了します。	1
	DOCFDELWRNMODE	添付ファイルからのテキスト抽出がエラーとなった場合，または DOCFATTACHED に 2 を指定している場合に，ワーニングの詳細情報を残すかどうか，および残すファイルを指定します。	• 0：詳細情報ファイルと一時ファイルを残しません。 • 1：詳細情報ファイルを残します。 • 2：詳細情報ファイルと一時ファイルを残します。	1
	DOCFEXTMAX	テキスト抽出サイズを KB 単位で指定します。	1 ～ 102,400（KB）の範囲で指定します。	5,120（KB）

セクション	エントリ	内容	設定値	デフォルト値
[DOCFLIBRARYMAP]	XXXX = XXXX,XXXX	ファイルの拡張子に応じて使用するテキスト抽出ライブラリを分ける場合に，対応付けを指定します。 テキスト抽出ライブラリは，複数指定できます。 複数指定した場合は，先に指定したテキスト抽出ライブラリでテキスト抽出を実行し，テキスト抽出に失敗した場合に次のテキスト抽出ライブラリでテキスト抽出を実行します。 記述形式： 拡張子※2＝ライブラリ種別 [, ライブラリ種別…] すべての拡張子，および拡張子なしを指定する場合は *（アスタリスク）を指定します。拡張子は大文字，小文字を区別しません。 * 指定と拡張子の指定の両方がある場合は，指定した拡張子は * 指定の対象外になります。	• DMC DMC を指定した場合は，DMC フィルタを適用します。 • MSIF MSIF を指定した場合は，32bit 版の IFilter を適用します。 • MS64 MS64 を指定した場合は，64bit 版の IFilter を適用します。 • DOCF DOCF を指定した場合は，DOCF フィルタを適用します。 • DMTX DMTX を指定した場合は，DMTX フィルタを適用します。	* = DMC txt = DMTX,DOCF
[DOCFTYPEMAP]	XXXX = XXXX	対応している文書種別と同じ扱いにしたい場合に，仮定拡張子を指定します。※5 記述形式： 元拡張子＝仮定拡張子 拡張子の対応をエントリ値に指定します。 抽出元の拡張子を元拡張子に，テキスト抽出ライブラリに渡す拡張子を仮定拡張子に指定します。	• 拡張子 元拡張子には，「.」（ピリオド）も指定できます。 ピリオドの場合は，拡張子なしの指定になります。	拡張子の扱いを変更しません。※2，※3

注※1

標準コンフィグレーションファイルでは，600 を設定しています。

注※2

拡張子は，大文字・小文字を区別しません。

注※ 3

拡張子および仮定拡張子は 32 バイトまで設定できます。

注※ 4

一時フォルダは，Document Filter for Text Search 専用に設定してください。

注※ 5

「1.3　対象ドキュメント」の表 1-1 に記載している文書種別を表す拡張子と異なる拡張子のファイルを対象とします。

2.3.2　記述規則

コンフィグレーションファイルの記述規則を次に示します。

- セクション名は [　] で囲んで記述します。
- 各エントリに値を設定するには，「エントリ名＝設定値」の形式で記述します。このとき，エントリ名と =，= と設定値の間にスペースまたはタブを挿入できます。
- 行の先頭または行中に #（シャープ）を記述すると，# 以降改行コードまでコメントとして扱われます。
- 同じエントリ名に対する設定を複数記述した場合は，最初に記述した内容が有効となります。
- 定義ファイルの最終行には，必ず改行コードを記述してください。

2.3.3　定義例

コンフィグレーションファイルの定義例を次の図に示します。

図 2-1　コンフィグレーションファイルの定義例

```
[SYSTEM]
  DOCFTIMEOUT     = 600        1つのファイルからのテキスト抽出が10分(600秒)を超えた
                               場合，対象ファイルのテキスト抽出を中止します。
  DOCFWORKDIR     = d:¥tmp     一時ファイルをd:¥tmpに作成します。
[DOCFLIBRARYMAP]
  TXT             = DMTX,DOCF  拡張子TXTのファイルを，DMTXでテキスト抽出します。
                               DMTXで抽出できなかったファイルは，DOCFで再度テキスト
                               抽出します。
  pdf             = DMC,MSIF   拡張子pdfのファイルを，DMCでキスト抽出します。DMCで
                               抽出できなかったファイルは，MSIFで再度テキスト抽出し
                               ます。
  *               = DMC        拡張子がTXTとpdf以外のファイルはすべてDMCでテキスト
                               抽出します。
[DOCFTYPEMAP]
  C               = TXT        拡張子cのファイルを，拡張子TXTとして扱います。
  CSV             = TXT        拡張子csvのファイルを，拡張子TXTとして扱います。
  .               = TXT        拡張子がないファイルを，拡張子TXTとして扱います。
```

2.4 テキスト抽出時の注意事項

テキスト抽出時の注意事項を次に示します。

● インストールについて

Document Filter for Text Search は，上位プログラムに対してテキスト抽出機能を提供するライブラリで構成されています。上位プログラムでテキスト抽出機能を利用する場合，Document Filter for Text Search を必ずインストールしてください。

● テキスト抽出できるファイルについて

Document Filter for Text Search は，表 2-2 に示す各種アプリケーションで作成したファイルからテキスト抽出できます。
サポートしているフィルタ，文書種類，使用するプロパティ情報，およびその扱い方は上位プログラムに依存します。詳細については，上位プログラムのマニュアルをご覧ください。

DMC フィルタの対象ドキュメントの種類を，次の表に示します。

なお，文書種類は各アプリケーションの日本語版に対応しています。また，プラットフォームによって表 2-2 に示す内容と異なる場合があります。詳細はリリースノートを参照してください。

表 2-2 対象ドキュメントの詳細［DMC フィルタ］

文書種類	バージョンおよび形式（拡張子※1）
Word	Word 95，Word 97，Word 98，Word 2000，Word 2002，Word 2003，Word 2007，Word2010
	Word 文書形式※2（DOC，DOT，DOCX，DOCM，DOTX，DOTM）
Excel	Excel 95，Excel 97，Excel 2000，Excel 2002，Excel 2003，Excel 2007，Excel 2010
	Excel ブック形式（XLS，XLT，XLSX，XLSM，XLTX，XLTM）
PowerPoint	PowerPoint 95，PowerPoint 97，PowerPoint 2000，PowerPoint 2002，PowerPoint 2003，PowerPoint 2007，PowerPoint 2010
	スライドショー形式，プレゼンテーション形式※2（PPT，POT，PPS，PPTX，PPTM，POTX，POTM，PPSX，PPSM）
一太郎	Version 7，Version 8，Version 9，Version 10，Version 11，Version 12，Version 13，2004，2005，2006，2007，2008，2009，2010，2011
	通常形式，7 形式（JTD，JTT，JFW，JVW）
PDF	PDF1.2，PDF1.3，PDF1.4，PDF1.5，PDF1.6，PDF1.7，ISO 32000-1 準拠
	Acrobat で作成した文書形式，ISO 32000-1 準拠の PDF ファイル形式（PDF）
Lotus 1-2-3	R5J，97，98，2000，Millennium Edition（WK4，123）

文書種類	バージョンおよび形式（拡張子※1）
OASYS	V5，V6，V7，V8，2002
	結合型ファイル形式，複合型ファイル形式（OA2，OA3）
DocuWorks	V4，V5.，V6
	DocuWorks 文書形式（XDW）
RTF	RTF 1.3，1.4，1.5
	Office の RTF 形式で出力した文書（RTF）
HTML	HTML タグを含むファイル（HTM，HTML）
XML	<?xml version="1.0" . . . . ?> があるファイル（XML）

注※ 1

本ライブラリは形式をバイナリ判定するため，拡張子には依存しません。アプリケーションが使用する拡張子です。

注※ 2

文書形式は Microsoft Office Binary File Formats，および Microsoft Office Binary File Formats に対応する Office Open XML File Formats です。

DMC フィルタの対象ドキュメントは，ファイル中に別のファイル形式を含む場合があります。含まれるファイル形式が添付ファイル，または OLE オブジェクトの場合に，テキスト抽出できる対象ドキュメントの文書形式を，表 2-3，表 2-4 に示します。

表 2-3　添付ファイルからテキスト抽出できる文書の形式

添付ファイル付き文書の種類と形式	テキスト抽出対象の添付ファイル
PDF • 添付ファイル形式 • PDF パッケージ形式 • PDF ポートフォリオ形式	• DMC フィルタの対象ドキュメント • テキストファイル
DocuWorks • オリジナルデータ添付形式	

表 2-4 OLE オブジェクトからテキスト抽出できる文書の形式

OLE オブジェクト付き文書の種類	テキスト抽出対象の OLE オブジェクト
• Word • Excel • PowerPoint • 一太郎	• Word • Excel • PowerPoint • RTF • PDF • 一太郎 • DocuWorks

MSIF/MS64 フィルタの対象ドキュメントの詳細を，次の表に示します。

表 2-5 対象ドキュメントの詳細［MSIF/MS64 フィルタ］

文書種類	バージョン
任意	導入する IFilter に依存します。
	導入する IFilter が対応する形式のファイル

DOCF フィルタの対象ドキュメントの詳細を，次の表に示します。

表 2-6 対象ドキュメントの詳細［DOCF フィルタ］

文書種類	バージョン
テキスト	Shift-JIS（Windows31J），EUC-JP，JIS，UTF-8（UCS-2 範囲）
	拡張子が txt のファイル

DMTX フィルタの対象ドキュメントの詳細を，次の表に示します。

表 2-7 対象ドキュメントの詳細［DMTX フィルタ］

文書種類	バージョン
テキスト	Shift-JIS（Windows31J），EUC-JP，JIS，UTF-8（UCS-4 範囲），UTF-16
	テキストエディタ（メモ帳など）で作成したファイル

注意事項

一時フォルダ内のファイルについて
Document Filter for Text Search を利用し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，一時ファイルの出力先に一時ファイルが残る場合があります。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の圧迫につながりますので，必要に応じて削除してください。

テキスト抽出する文書ファイル，および使用するフィルタの注意事項を次に示します。

2.4.1 DMC フィルタの場合

(1) 共通の注意事項

- テキスト抽出サイズをデフォルト（5MB）以上にする場合は，上位プログラム，および組み込まれているシステムがサポートしているサイズ以内に設定してください。
- 添付ファイルの表示順序と，添付ファイルからテキスト抽出する順序は一致しない場合があります。
- 添付ファイル，および OLE オブジェクトの場合は，その文書種別（拡張子）がコンフィグレーションの設定で別のテキスト抽出ライブラリの適用を定義されているときでも，DMC フィルタでテキスト抽出されます。
  また，テキストファイルの判定，抽出結果は DMTX フィルタを使用した場合と同じです。コンフィグレーションの定義は適用されません。

次の文書ファイル，および文書の情報はテキスト抽出できない場合があります。

- 読み出しのパスワードが設定されている文書ファイル
- 図形，イメージ，線画，枠，数式で構成されている文書ファイル
- ヘッダー，フッター，ページ番号，および番号などの自動で生成する情報
- Microsoft の IRM（Information Right Management）機能を使用し，文書へのアクセス権限（閲覧・変更など）を設定した Word 2003，Excel 2003，PowerPoint2003 の文書ファイル
- 実行中のアプリケーションと異なる形式やバージョンで保存した文書ファイル
- レイアウト表示どおりにテキストが格納されていない文書ファイル
- リンク形式の OLE オブジェクト
- 4 階層以上の添付ファイル，または OLE オブジェクトを含む文書ファイル
- OLE オブジェクトの作成方法によっては，テキスト抽出できません

次の場合は，詳細情報が出力されます。

- 添付ファイルをテキスト抽出できない文書と判断した場合
  ただし，未サポートの添付ファイルの場合は詳細情報は出力されません。

(2) Word の注意事項

- 自動更新の日付，時間は正しくテキスト抽出できません。
- 特殊文字の一部は，テキスト抽出できません。
- フィールドの内容は，一部テキスト抽出できません。
- 日本語環境以外でローカライズされた Word により作成されたファイルは，テキスト抽出できない場合があります。

(3) Excel の注意事項

- ヘッダー，フッターでは，指定されたページ番号，頁数，日付，時刻，ファイル名，シート名はテキスト抽出できません。
- 「シートの保護」を設定したファイルのテキスト抽出はできますが，「ブックの保護」を指定したファイルはテキスト抽出できません。
- 「ユーザ定義」のセルは，表示結果とテキスト抽出の結果が異なる場合があります。
- 日本語環境以外でローカライズされた Excel により作成されたファイルは，テキスト抽出できない場合があります。

(4) PowerPoint の注意事項

スライドとノート以外はテキスト抽出できません。

(5) 一太郎の注意事項

- 圧縮してから保存した文書はテキスト抽出できません。
- マスキング文書の塗りつぶされた枠内はテキスト抽出できません。

(6) PDF の注意事項

- ユーザ定義 Cmap 文字は抽出できない場合があります。また，Symbol 文字は文字化けする場合があります。
- 立体文字は，複数の文字がテキスト抽出されます。
- 文字のフォントが Wingdings の場合は，該当文字が抽出できません。例えば，Word 文書，PowerPoint 文書で作成した箇条書きの行頭文字（○●◆□■など）を PDF 文書に変換した場合，該当の行頭文字は正しく抽出できません。
- Acrobat の「テキスト選択ツール」でコピーできない文字は，抽出できません。
- サポートしているバージョン以外の Acrobat や，別のアプリケーションで作成した PDF は，テキスト抽出に失敗したり，文字化けしたりする場合があります。
- ユーザ定義文字や PDF 独自のコードが使用されている文字は，文字化けする場合があります。
- 文書を開くパスワードが設定されている文書は，テキスト抽出できません。
- OwnerPassWord の設定されている文書は，40-Bit RC4 の場合を除きテキスト抽出できません。
- パスワードによるセキュリティ設定で，テキスト内容のコピー（抽出）が許可されていない場合はテキスト抽出できません。
  テキスト内容のコピーが許可されている場合は，その他のセキュリティ権限（変更・印刷等）が許可されていなくてもテキスト抽出できます。

- Acrobat のセキュリティの選択で，互換性のある形式が「Acrobat 7 およびそれ以降」に設定されている文書からはテキスト抽出できません。
- TYPE3 フォントは抽出できません。
- パッケージ形式，およびポートフォリオ形式の場合は，表紙（テンプレート）からもテキスト抽出されます。
- ポートフォリオ形式のタイトル，およびカバーからはテキスト抽出できません。

（7） DocuWorks の注意事項

- 太文字，影付きで修飾された文字は，複数の文字がテキスト抽出されます。
- セキュリティが設定された文書はテキスト抽出できません。
- 縦書きテキストは一文字ごとに改行されます。このため，単語として扱えない場合があります。
- OLE オブジェクトからは，テキスト抽出できません。
- 署名された文書は，セキュリティが設定された文書として扱われるためテキスト抽出できません。

（8） RTF の注意事項

- 自動更新の日付，時間は正しくテキスト抽出できません。
- 特殊文字の一部は，テキスト抽出できません。
- フィールドの内容の一部は，テキスト抽出できません。
- 日本語環境以外でローカライズされたアプリケーションにより作成されたファイルは，テキスト抽出できない場合があります。

（9） HTML の注意事項

- タグと属性の内容はテキスト抽出できません。
- <html> タグが無い場合は，テキスト抽出できません。
- META タグに文字コードセット指定が無い場合に，EUC コードで記述された文書ファイルはテキスト抽出できません。

（10） XML の注意事項

Microsoft Office で作成した XML ファイルに OLE オブジェクトが存在する場合は，OLE オブジェクトの抽出はできません。

2.4.2 MSIF/MS64 フィルタの場合

- MSIF/MS64 フィルタは拡張子のマッピングによって，処理できるファイルが決定されます。このため上位アプリケーションの機能によっては使用できない場合があります。対象ファイルの拡張子が異なる場合は，拡張子を変更するかコンフィグレーションの「DOCFTYPEMAP」を使用してください。
- 64bit OS では適用可能な IFilter が 64bit の場合と 32bit（WOW64 モード）の場合があります。適用した IFilter のアーキテクチャに従ったフィルタを使用してください。

2.4.3 DOCF フィルタの場合

- 処理できるファイルの拡張子は，txt だけです。
- 対象ファイルの拡張子が異なる場合は，拡張子を変更するかコンフィグレーションの「DOCFTYPEMAP」を使用してください。
- コード種別を自動判定する場合，含まれるコード範囲によっては正しく判断できないときがあります。判定できなかったファイルは，コード変換を行わずそのまま抽出結果とします。

2.4.4 DMTX フィルタの場合

テキストファイルに含まれるコード範囲によっては，テキストファイルと判断できない場合があります。

上位プログラムがテキスト抽出に失敗した場合に，メッセージに Document Filter for Text Search の詳細コードを示すときがあります。詳細コードの詳細については，「3.4 詳細コード」を参照してください。

3 障害対策

Document Filter for Text Search で障害が発生した場合，上位アプリケーションの障害対策にしたがって対策してください。この章では，Document Filter for Text Search の障害発生時に出力される詳細情報ファイルおよび一時ファイルについて説明します。

3.1 **障害情報の取得**

3.2 **詳細情報ファイル**

3.3 **一時ファイル**

3.4 **詳細コード**

3.1 障害情報の取得

Document Filter for Text Search の上位アプリケーションの情報，および「3.4 詳細コード」で，対処方法が不明の場合は，障害情報を取得する必要があります。

障害情報を取得する場合はシステム管理者の指示にしたがって，「3.2 詳細情報ファイル」「3.3 一時ファイル」の障害情報ファイルを取得してください。

障害情報ファイルの出力先や保管方法は，コンフィグレーションファイルの設定によって変わります。詳細は，「2.3 コンフィグレーションファイルの設定」を参照してください。

3.2 詳細情報ファイル

Document Filter for Text Search に障害が発生した場合，関数の戻り値や，関数によって取得されるメッセージテキストが，詳細情報ファイルに出力されます。また，詳細情報ファイルが出力されるときに，一時ファイルが出力されます。

この節では，詳細情報ファイルの出力先，ファイル名の形式，出力形式，および出力例について説明します。

出力先

＜Document Filter for Text Searchのインストールフォルダ＞¥spool

ファイル名の形式

[time]_[pid].log

time：カレンダ時間[ミリ秒]（13けたの数字文字列）
pid：実行コマンドのプロセスID（10けたの数字文字列）

出力形式

YYYY/MM/DD hh:mm:ss.000 詳細情報（パラメタ情報，実行されたテキスト抽出ライブラリの関数情報など）

出力例

```
2008/10/07 11:31:22.184 start : docfdmc.exe
                              :
2008/10/07 11:31:22.934 DMC dflCmFileoutOfText DMC_GetText_V3
3004
2008/10/07 11:31:22.981 end : docfdmc.exe    27※
```

※注

終了行の終了コードが 27 の場合は，添付ファイルなど一部分のテキスト抽出ができなかったことを意味します。

詳細情報の保管方法をコンフィグレーションファイルに設定できます。詳細情報の保管方法の設定については，「2.3　コンフィグレーションファイルの設定」を参照してください。

3.3 一時ファイル

詳細情報ファイルが出力されるときに，一時ファイルが出力されます。この節では，一時ファイルの出力先について説明します。

一時ファイルの構成

出力先

＜Document Filter for Text Searchのインストールフォルダ＞¥tmp¥セッションID

セッション ID

docf+23けたの数字文字列
例）docf12345678901231234567890
本ライブラリを使用するプログラムの抽出セッションごとに，ディレクトリを作成します。

一時ファイルの出力先をコンフィグレーションファイルに設定できます。一時フォルダの設定については，「2.3　コンフィグレーションファイルの設定」を参照してください。

3.4 詳細コード

Document Filter for Text Search で障害が発生した場合，上位プログラムに詳細コードを返却します。

この節では，上位プログラムのメッセージに本製品の詳細コードを参照するように指示がある場合の詳細コードと対処方法について説明します。

表 3-1 に記載されていない詳細コードの場合，システム管理者に連絡してください。

表 3-1 詳細コードと対処方法

値	対処方法
1	ファイルシステムを確認し，再実行してください。
2	システムで使用しているメモリを確認し，不要なプログラムを終了，または使用可能なメモリを増加して再実行してください。
3	ファイルシステムを確認し，再実行してください。
4	コンフィグレーションファイルの内容を確認して再実行してください。
5	空ファイルが指定されました。
6	パスワードが設定されているファイルが指定されました。
7	ファイルシステムまたはコンフィグレーションファイルの内容を確認して再実行してください。
8	未サポートの文書が指定されました。
	パスワードが設定されているファイルが指定されました。
9	扱えない文書が指定されました。
10	コンフィグレーションファイルのタイムアウト設定または上位プログラムのタイムアウト設定を変更して，再実行してください。
13	コンフィグレーションファイルの内容を確認して再実行してください。
14	コンフィグレーションファイルの内容を確認して再実行してください。
16	ファイルシステムを確認して再実行してください。
17	コンフィグレーションファイルの内容を確認して再実行してください。
18	ファイルシステムまたはコンフィグレーションファイルの内容を確認して再実行してください。
20	未サポートのローカライズの文書が指定されました。

表 3-2 表 3-2Document Filter for Text Search 02-30 以降を前提としていない上位プログラムの場合

値	対処方法
4	未サポートの文書が指定されました。
	未サポートのローカライズの文書が指定されました。
7	ファイルシステムまたはコンフィグレーションファイルを確認して再実行してください。

値	対処方法
10	空のファイルが指定されました。
11	パスワードが設定されているファイルが指定されました。
30	システムで使用しているメモリを確認し，不要なプログラムを終了，または使用可能なメモリを増加して再実行してください。
32	テキスト抽出処理がタイムアウトしました。
	コンフィグレーションファイルの指定に誤りがあります。次の事項を確認してください。 • 各エントリの指定が改行で終了しているか • DOCFWORKDIR エントリで指定している一時フォルダのパス名が 149 バイト以内か
	テキスト抽出処理が異常終了しました。システム管理者へ連絡してください。

付録

付録 A　フォルダ構成

付録 B　このマニュアルの参考情報

付録 C　用語解説

付録 A　フォルダ構成

Document Filter for Text Search のフォルダ構成を次に示します。

表 A-1　Document Filter for Text Search のフォルダ構成

フォルダ名	内容
インストールフォルダ /	• Windows の場合 < インストール時のユーザ指定フォルダ > • AIX，HP-UX または Solaris の場合 /opt/DocFilterTS インストールフォルダ以降の構成は，OS 共通です。
/config/	コンフィグレーションファイル格納フォルダ
config.cfg	テキスト抽出方法を定義したコンフィグレーションファイル
/spool/	詳細情報ファイル出力フォルダ 詳細情報ファイルについては，「3.2　詳細情報ファイル」を参照してください。
/tmp/	一時ファイル出力フォルダ 各テキスト抽出ライブラリで一時ファイルの作成が必要な場合に使用します。 一時ファイルについては，「3.3　一時ファイル」を参照してください。
/sample/	コンフィグレーションファイルのサンプル格納フォルダ
config.cfg	標準コンフィグレーションファイル

付録 B　このマニュアルの参考情報

このマニュアルを読むにあたっての参考情報を示します。

付録 B.1　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。

- DocumentBroker Text Search Index Loader Version 2（3000-3-790）
- uCosminexs DocumentBroker Text Search Index Loader Version 3（3020-3-N46）
- Groupmax Document Manager Version 6 システム管理者ガイド（3020-3-B54）
- HiRDB Text Search Plug-in Index Generator（3000-6-289）
- uCosminexus Enterprise Search 環境設定ガイド（3020-3-H90）
- uCosminexus Enterprise Search 運用ガイド（3020-3-H91）

付録 B.2　このマニュアルでの表記

このマニュアルでは，製品名を次のように表記しています。

製品名称	表記
AIX 5L V5.2	AIX
AIX 5L V5.3	
AIX V6.1	
DocumentBroker Development Kit Version 2	DocumentBroker
DocumentBroker Repository Version 2	
DocumentBroker Runtime Version 2	
DocumentBroker Server Version 2	
DocumentBroker Web Client Version 2	
DocumentBroker Text Search Index Loader Version 2	DocumentBroker Text Search Index Loader
uCosminexs DocumentBroker Text Search Index Loader Version 3	
Document Filter for Text Search Version 3	Document Filter for Text Search
Groupmax Document Manager Version 6	Groupmax Document Manager
HP-UX 11，11i	HP-UX
Solaris 8	Solaris
Solaris 9	
Solaris 10	

製品名称	表記
uCosminexus Enterprise Search	Enterprise Search

付録 B.3　英略語

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	説明
API	Application Programming Interface
CSV	Comma Separated Values
DLL	Dynamic Linking Library
EUC	Extended Unix Code
HTML	Hyper Text Markup Language
ISO	International Organization for Standardization
JIS	Japanese Industrial Standards
OLE	Object Linking and Embedding
OS	Operating System
PDF	Portable Document Format
PC	Personal Computer
RTF	Rich Text Format
WOW64	Windows 32-bit On Windows 64-bit
XML	eXtensible Markup Language

付録 B.4　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。

付録 C　用語解説

（英字）

OLE オブジェクト付き文書

文書の種類が，Word，Excel，PowerPoint，一太郎の場合，文書内に OLE オブジェクトを埋め込むことができます。Document Filter for Text Search では，文書内に OLE オブジェクトが存在する文書をまとめて，OLE オブジェクト付き文書と呼びます。

（サ行）

詳細情報ファイル

Document Filter for Text Search に障害が発生した場合，関数の戻り値や，関数によって取得されるメッセージテキストを出力するログファイルのことです。

（タ行）

抽出結果文書

Document Filter for Text Search が抽出したテキストデータのことです。

抽出元文書

テキストデータを抽出する前のデータのことです。

テキストデータ

文書の構成要素の一つです。テキストデータは，属性以外のテキスト情報を指します。

添付ファイル付き文書

文書の種類が PDF，または DocuWorks の場合，文書形式の種類に添付ファイル形式があります。Document Filter for Text Search では，添付ファイル形式の PDF，DocWorks をまとめて，添付ファイル付き文書と呼びます。

索引

C

config.cfg 15

D

DMC フィルタの場合 22
DMTX フィルタの場合 25
DOCF フィルタの場合 25
Document Filter for Text Search とは 2

M

MSIF/MS64 フィルタの場合 24

O

OLE オブジェクト付き文書 37

あ

アンインストールの方法 12

い

一時ファイル 30
インストールとアンインストール 9
インストールの方法 9

か

概要 1
環境設定 7
環境設定の流れ 8

き

記述規則 18

こ

コンフィグレーションファイルの設定 15

し

システム構成 2
障害情報の取得 28
障害対策 27
詳細コード 31
詳細情報ファイル 29,37
処理概要 3

た

対象ドキュメント 4

ち

抽出結果文書 37
抽出するプロパティ情報 5
抽出元文書 37

て

定義内容 15
定義例 18
テキスト抽出時の注意事項 19
テキストデータ 37
添付ファイル付き文書 37

ふ

フォルダ構成 34
文書種別 4,5