

---

# HiRDB Text Search Plug-in Index Generator

解説書

3000-6-289

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，  
巻末の用紙をご利用ください。

**HITACHI**

## 対象製品

P-24D3-3914 HiRDB Text Search Plug-in Index Generator 01-00 (適用 OS : Windows 2000 , Windows Server 2003 )

P-1MD3-3911 HiRDB Text Search Plug-in Index Generator 01-00 (適用 OS : AIX 5L V5.1 , AIX 5L V5.2 )

これら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「ソフトウェア添付資料」または「Readme ファイル」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

AIX は、米国における米国 International Business Machines Corp. 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。

UNIX は、X/Open Company Limited が独占的にライセンスしている米国ならびに他の国における登録商標です。

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

X/Open は、X/Open Company Limited の英国ならびに他の国における登録商標です。

## 発行

2004 年 3 月 (第 1 版) 3000-6-289

## 著作権

All Rights Reserved. Copyright (C) 2004, Hitachi, Ltd.

# はじめに

---

このマニュアルは HiRDB Text Search Plug-in Index Generator の機能および利用方法について説明したものです。

P-24D3-3914 HiRDB Text Search Plug-in Index Generator

P-1MD3-3911 HiRDB Text Search Plug-in Index Generator

以降、このマニュアルでは、上記のプログラムプロダクトの総称を Text Search Plug-in Index Generator とします。

## 対象読者

Text Search Plug-in Index Generator を使って、電子ドキュメントから全文検索用テキストデータを抽出したい方を対象としています。また、次の知識を持っていることを前提としています。

- Windows または UNIX に関する基本的な操作方法を知っている方
- HiRDB に関する知識および操作方法を知っている方

## マニュアルの構成

このマニュアルは、次に示す五つの章と付録から構成されています。

### 第 1 章 概要

Text Search Plug-in Index Generator の機能概要およびシステム構成について説明しています。

### 第 2 章 環境設定

Text Search Plug-in Index Generator を利用するための、環境設定方法について説明しています。

### 第 3 章 全文検索用データの登録

Text Search Plug-in Index Generator を使って、全文検索用データをデータベースに登録するためのコマンドの機能、およびコマンドの処理の定義方法について説明しています。

### 第 4 章 運用

Text Search Plug-in Index Generator を効率良く運用する方法について説明しています。

### 第 5 章 障害対策

コマンド実行時に発生した障害の対処方法、およびメッセージとその対処方法について説明しています。

### 付録 A ディレクトリ構成

Text Search Plug-in Index Generator のディレクトリ構成を示しています。

### 付録 B 用語解説

このマニュアルで使用する用語のうち、説明の必要な用語について解説しています。

## 関連マニュアル

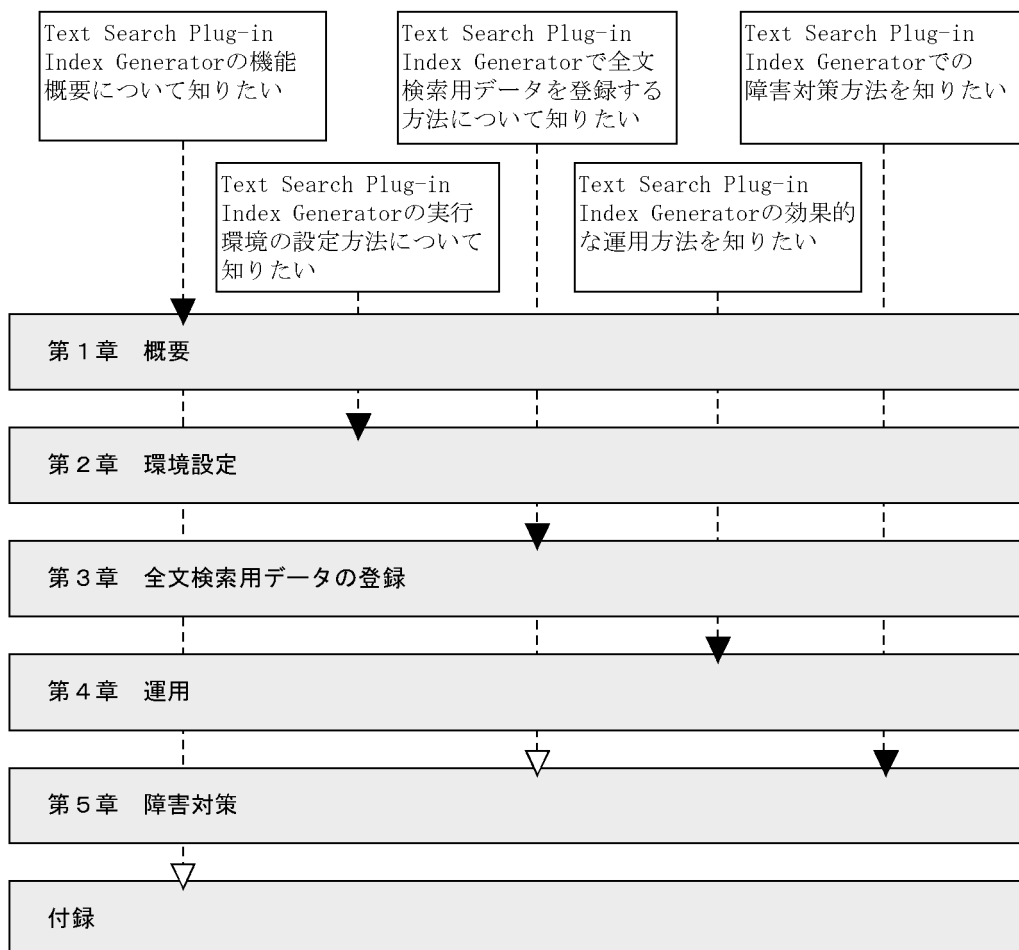
このマニュアルの関連マニュアルを次に示します。必要に応じてお読みください。

- HiRDB Text Search Plug-in Version 7 ( 3000-6-288 )
- HiRDB Version 6 解説 ( UNIX(R) 用 )( 3000-6-231 )
- HiRDB Version 7 解説 ( UNIX(R) 用 )( 3000-6-271 )
- HiRDB Version 6 システム導入・設計ガイド ( UNIX(R) 用 )( 3000-6-232 )
- HiRDB Version 7 システム導入・設計ガイド ( UNIX(R) 用 )( 3000-6-272 )
- HiRDB Version 6 システム定義 ( UNIX(R) 用 )( 3000-6-233 )
- HiRDB Version 7 システム定義 ( UNIX(R) 用 )( 3000-6-273 )
- HiRDB Version 6 システム運用ガイド ( UNIX(R) 用 )( 3000-6-234 )
- HiRDB Version 7 システム運用ガイド ( UNIX(R) 用 )( 3000-6-274 )
- HiRDB Version 6 コマンドリファレンス ( UNIX(R) 用 )( 3000-6-235 )
- HiRDB Version 7 コマンドリファレンス ( UNIX(R) 用 )( 3000-6-275 )
- HiRDB Version 6 UAP 開発ガイド ( UNIX(R)/Windows(R) 用 )( 3000-6-236 )
- HiRDB Version 7 UAP 開発ガイド ( UNIX(R)/Windows(R) 用 )( 3000-6-276 )
- HiRDB Version 6 SQL リファレンス ( UNIX(R)/Windows(R) 用 )( 3000-6-237 )
- HiRDB Version 7 SQL リファレンス ( UNIX(R)/Windows(R) 用 )( 3000-6-277 )
- HiRDB Version 6 メッセージ ( UNIX(R)/Windows(R) 用 )( 3000-6-238 )
- HiRDB Version 7 メッセージ ( UNIX(R)/Windows(R) 用 )( 3000-6-278 )
- HiRDB Version 6 解説 ( Windows(R) 用 )( 3020-6-121 )
- HiRDB Version 7 解説 ( Windows(R) 用 )( 3020-6-271 )
- HiRDB Version 6 システム導入・設計ガイド ( Windows(R) 用 )( 3020-6-122 )
- HiRDB Version 7 システム導入・設計ガイド ( Windows(R) 用 )( 3020-6-272 )
- HiRDB Version 6 システム定義 ( Windows(R) 用 )( 3020-6-123 )
- HiRDB Version 7 システム定義 ( Windows(R) 用 )( 3020-6-273 )
- HiRDB Version 6 システム運用ガイド ( Windows(R) 用 )( 3020-6-124 )
- HiRDB Version 7 システム運用ガイド ( Windows(R) 用 )( 3020-6-274 )
- HiRDB Version 6 コマンドリファレンス ( Windows(R) 用 )( 3020-6-125 )
- HiRDB Version 7 コマンドリファレンス ( Windows(R) 用 )( 3020-6-275 )
- HiRDB Version 6 UAP 開発ガイド ( Windows(R) 用 )( 3020-6-126 )
- HiRDB Version 7 UAP 開発ガイド ( Windows(R) 用 )( 3020-6-276 )
- HiRDB Version 6 SQL リファレンス ( Windows(R) 用 )( 3020-6-127 )
- HiRDB Version 7 SQL リファレンス ( Windows(R) 用 )( 3020-6-277 )
- HiRDB Version 6 メッセージ ( Windows(R) 用 )( 3020-6-128 )
- HiRDB Version 7 メッセージ ( Windows(R) 用 )( 3020-6-278 )
- Document Filter for Text Search Version 3 ( 3020-3-D59 )

なお、このマニュアルでは、HiRDB 関連のマニュアルについて「( UNIX(R) 用 )」および「( UNIX(R)/Windows(R) 用 )」および「( Windows(R) 用 )」を省略して表記しています。また、「HiRDB Version 6」と「HiRDB Version 7」を合わせて「HiRDB」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読むことができます。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

## このマニュアルの図中で使用する記号

このマニュアルの図中で使用する記号を、次のように定義します。

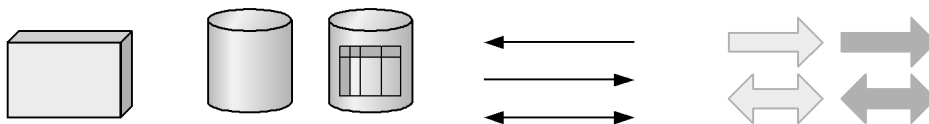
はじめに

●プログラム

●データベース

●その他の流れ

●データの流れ



## こ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主な英略語を次に示します。

英略語	説明
AP	Application Program
BLOB	Binary Large Object
DB	Data Base
JIS	Japanese Industrial Standard code
OS	Operating System
RDB	Relational Data Base
SGML	Standard Generalized Markup Language
SQL	Structured Query Language
UAP	User Application Program

## このマニュアルでの表記

このマニュアルで使用する製品名称の略称を次に示します。

製品名称	略称	
Document Filter for Text Search Version 3	Document Filter for Text Search	
HiRDB/Single Server Version 7	HiRDB/ シングルサーバ (またはシングルサーバ)	HiRDB
HiRDB/Parallel Server Version 7	HiRDB/ パラレルサーバ (またはパラレルサーバ)	
HiRDB/Developer's Kit Version 7 および HiRDB/Run Time Version 7	HiRDB クライアント	
HiRDB/Single Server・Object Option および HiRDB/Parallel Server・Object Option	Object Option	
HiRDB Text Search Plug-in Version 7	Text Search Plug-in	
HiRDB Text Search Plug-in Conceptual Extension Version 7	Text Search Plug-in Conceptual Extension	

製品名称	略称	
Microsoft(R) Windows 2000(R) Professional Operating System, Microsoft(R) Windows 2000(R) Server Operating System, または Microsoft(R) Windows 2000(R) Advanced Server Operating System	Windows 2000	Windows
Microsoft(R) Windows Server(TM) 2003, Standard Edition	Windows Server 2003	
Microsoft(R) Windows Server(TM) 2003, Enterprise Edition		

- このマニュアルでは、「ディレクトリ」と「フォルダ」は、「ディレクトリ」に統一して表記しています。また、パス名の区切りは「/」で表記している個所があります。Windows 版の Text Search Plug-in をご使用の場合は、マニュアル中の「/」を「¥」に置き換えてください。

### 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。

必須（ひつす）

### 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、1,024<sup>2</sup> バイト、1,024<sup>3</sup> バイト、1,024<sup>4</sup> バイトです。





# 目次

<b>1</b>	<b>概要</b>	<b>1</b>
1.1	Text Search Plug-in Index Generator の機能	2
1.2	システム構成	4
<b>2</b>	<b>環境設定</b>	<b>7</b>
2.1	実行環境の設定	8
2.1.1	前提プログラムの環境設定	8
2.1.2	環境変数の設定	8
2.2	入力表および出力表の構成	10
2.3	インデクスの作成	15
<b>3</b>	<b>全文検索用データの登録</b>	<b>17</b>
3.1	TSPIndexer (全文検索用データ登録コマンド)	18
3.2	制御情報ファイルの定義	20
3.2.1	制御情報ファイルの記述形式と規則	20
3.2.2	制御情報ファイルの指定項目と指定方法	20
3.2.3	制御情報ファイルの定義例	26
<b>4</b>	<b>運用</b>	<b>29</b>
4.1	TSPIndexer コマンド実行時の留意点	30
4.2	推奨する運用方法	31
<b>5</b>	<b>障害対策</b>	<b>33</b>
5.1	コマンド実行時の障害対策	34
5.1.1	障害の切り分け方法	34
5.1.2	障害発生時の対策方法	34
5.2	メッセージ	36
5.2.1	メッセージの形式と見方	36
5.2.2	メッセージ一覧	37

## 付録 47

---

付録 A ディレクトリ構成 48

---

付録 B 用語解説 49

---

## 索引 51

---

## 目次

図 1-1	Text Search Plug-in Index Generator の処理概要	3
図 1-2	Text Search Plug-in Index Generator のシステム構成	4
図 A-1	Text Search Plug-in Index Generator のディレクトリ構成 (Windows の場合)	48
図 A-2	Text Search Plug-in Index Generator のディレクトリ構成 (UNIX の場合)	48

## 表目次

表 2-1	全文検索用列が入力表と同じである場合の表構成	10
表 2-2	全文検索用列が入力表と異なる場合の入力表構成	11
表 2-3	全文検索用列が入力情報表と異なる場合の出力表の構成	11
表 3-1	制御情報ファイルに指定する項目	20
表 4-1	Text Search Plug-in で推奨するパラメタの組み合わせ	31

# 1

## 概要

この章では、Text Search Plug-in Index Generator の機能概要およびシステム構成について説明します。

---

1.1 Text Search Plug-in Index Generator の機能

---

1.2 システム構成

---

## 1.1 Text Search Plug-in Index Generator の機能

---

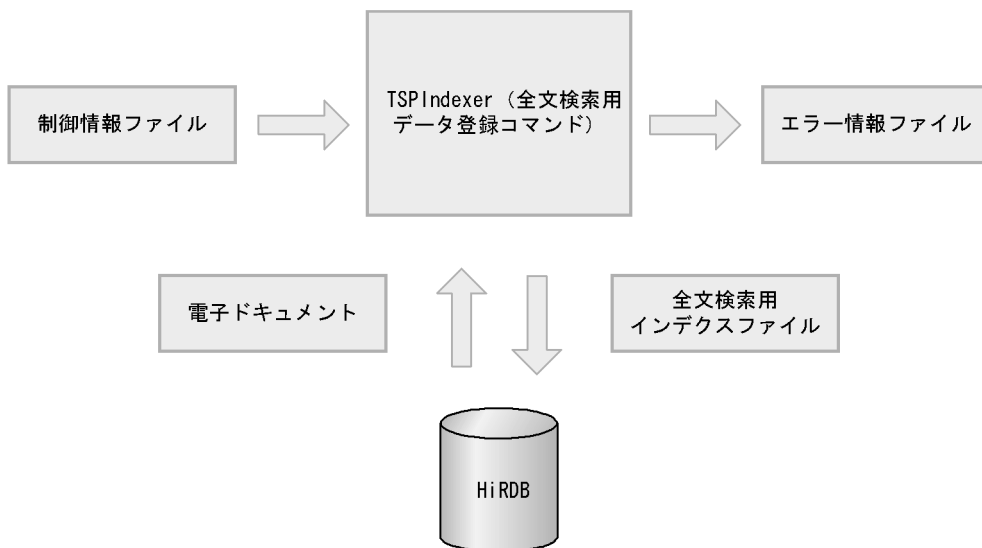
HiRDB を利用したシステムでは、Text Search Plug-in を使った高度な全文検索機能を利用できます。Text Search Plug-in の全文検索機能を利用する場合、検索用のテキストデータを HiRDB に登録する必要があります。企業や官公庁に限らず、一般に文書はワードプロセッサなどの文書作成ソフトウェアで作成される電子ドキュメントがほとんどです。このため、電子ドキュメントからテキストデータを抽出し、HiRDB に登録する仕組みを作成する必要があります。

Text Search Plug-in Index Generator は、HiRDB サーバの BLOB 型または BINARY 型列に格納されている電子ドキュメントからテキストを抽出し、全文検索用データとして SGMLTEXT 型または FREEWORD 型列にインデクス登録するプログラムです。Text Search Plug-in Index Generator を利用することで、文書作成ソフトウェアで作成した文書からテキストデータを抽出し、HiRDB に登録するユーザプログラムを作成する手間を省略できます。

Text Search Plug-in Index Generator が処理できる文書は、テキスト抽出に利用する Document Filter for Text Search がサポートしている文書です。詳細については、マニュアル「Document Filter for Text Search Version 3」を参照してください。

Text Search Plug-in Index Generator は、コマンドを実行することで文書を処理します。このコマンドを TSPIndexer コマンド（全文検索用データ登録コマンド）といいます。TSPIndexer コマンドは、実行条件を定義したファイル（制御情報ファイル）に基づいて動作します。Text Search Plug-in Index Generator の処理概要を次の図に示します。

図 1-1 Text Search Plug-in Index Generator の処理概要

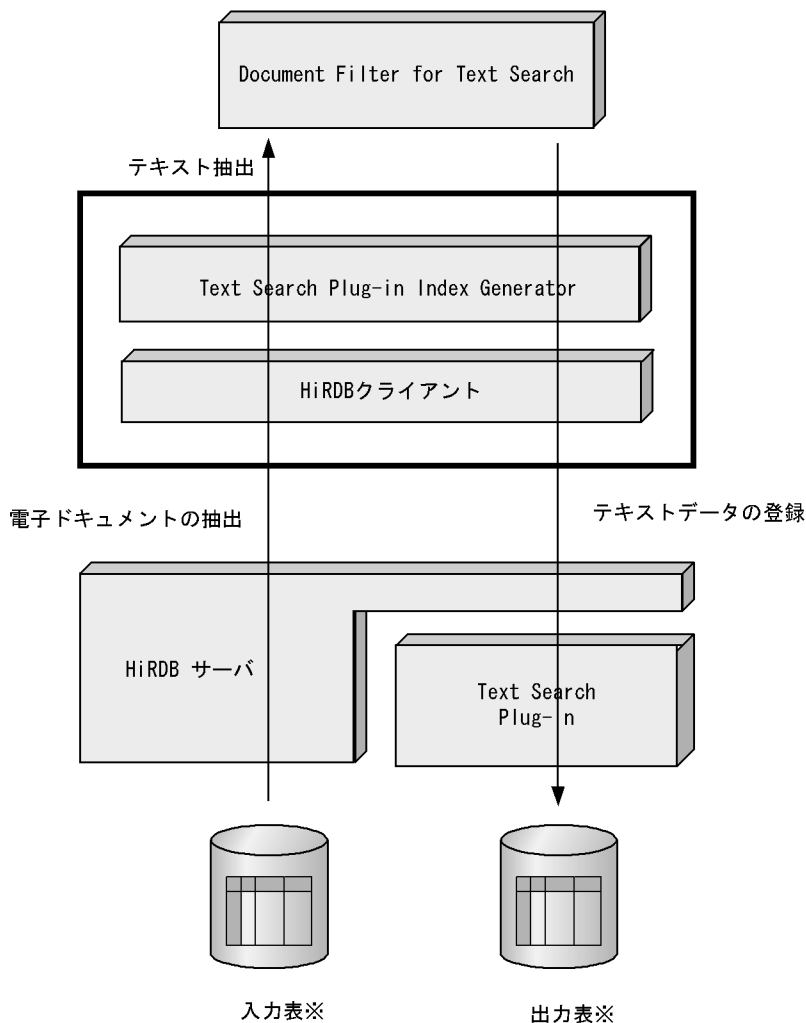


- 制御情報ファイル  
TSPIndexer コマンドの実行に必要な情報を定義します。
- エラー情報ファイル  
TSPIndexer コマンド実行時に発生したメッセージがエラーログの取得を指定した場合に、エラー情報を出力するファイルです。

## 1.2 システム構成

Text Search Plug-in Index Generator のシステム構成を次の図に示します。

図 1-2 Text Search Plug-in Index Generator のシステム構成



注※ 入力表と出力表は同じ表でもかまいません。

### Document Filter for Text Search

電子ドキュメントからテキストデータを抽出するプログラムです。Text Search Plug-in Index Generator が処理できる文書は、Document Filter for Text Search がサポートしている文書が対象となります。



### HiRDB サーバ

HiRDB/ シングルサーバまたは HiRDB/ パラレルサーバで構成される , データベースを配置するサーバ側システムのことです。

次のどちらかのプログラムが必要です。

- HiRDB/Parallel Server Version 6 または HiRDB/Parallel Server Version 7
- HiRDB/Single Server Version 6 または HiRDB/Single Server Version 7

### HiRDB クライアント

UAP を開発および実行するクライアント側システムのことです。

次のどちらかのプログラムが必要です。

- HiRDB/Run Time Version 6 または HiRDB/Run Time Version 7
- HiRDB/Developer's Kit Version 6 または HiRDB/Developer's Kit Version 7

### Text Search Plug-in

HiRDB に登録されている電子ドキュメントの全文検索を実現するプログラムです。

### Text Search Plug-in Index Generator

HiRDB に登録されている電子ドキュメントからテキストデータを取り出し , 全文検索用列に登録するプログラムです。

Text Search Plug-in Index Generator は , HiRDB サーバのクライアントアプリケーションとして動作します。このため , アプリケーションプログラムと同様 , HiRDB クライアントの環境定義や Text Search Plug-in の設定が有効となります。



# 2

## 環境設定

この章では、Text Search Plug-in Index Generator を利用するための実行環境の設定方法について説明します。また、Text Search Plug-in Index Generator の入力表および出力表の構成についても説明します。

---

2.1 実行環境の設定

---

2.2 入力表および出力表の構成

---

2.3 インデクスの作成

---

## 2.1 実行環境の設定

---

この節では、Text Search Plug-in Index Generator を利用するための環境設定について説明します。

### 2.1.1 前提プログラムの環境設定

Text Search Plug-in Index Generator を利用するには、次のプログラムが必要です。

- Document Filter for Text Search 03-00 以降
- HiRDB/Run Time Version 6 06-01 以降
- HiRDB/Developer's Kit Version 6 06-01 以降
- HiRDB/Parallel Server Version 6 06-01 以降

Document Filter for Text Search の環境設定については、マニュアル「Document Filter for Text Search Version 3」を参照してください。HiRDB の環境設定については、マニュアル「HiRDB システム導入・設計ガイド」を参照してください。

文字コード種別に関する注意

- HiRDB サーバのセットアップ時の文字コード種別は、SJIS または Unicode (utf-8) のどちらかにしてください。
- Text Search Plug-in の文字コード種別が Unicode (utf-8) の場合には、Text Search Plug-in Index Generator では MS-Unicode の使用が前提となります。

### 2.1.2 環境変数の設定

Text Search Plug-in Index Generator を使用する場合、Text Search Plug-in をシステムにインストールします。Text Search Plug-in の環境設定方法については、マニュアル「HiRDB Text Search Plug-in Version 2」または「HiRDB Text Search Plug-in Version 7」を参照してください。

通常のクライアントに必要な環境を定義してください。このとき、次の環境変数を指定してください。

PDUSER

HiRDB に接続するときに必要な、ユーザ ID とパスワードを指定します。この指定は省略できません。ただし、この指定を実行するには、入力表および出力表に対する参照・更新権限を持っている必要があります。

PDDBLOG

データベースの更新ログ取得方式を指定します。PDDBLOG の詳細については、マニュアル「HiRDB 解説」を参照してください。

PATH (AIX の場合は LIBPATH)

Document Filter for Text Search と HiRDB クライアントのライブラリのパスを指定します。Document Filter for Text Search のパスの設定については、マニュアル「Document Filter for Text Search Version 3」を参照してください。HiRDB クライアントのパスの設定については、マニュアル「HiRDB システム導入・設計ガイド」を参照してください。

#### PDBLKF に関するご注意

HiRDB のブロック転送の行数指定は、TSPIndexer コマンド実行時に最適な値を設定しています。PDBLKF 環境変数を設定していても、TSPIndexer コマンド実行時には TSPIndexer コマンドが設定した値が有効となります。

## 2.2 入力表および出力表の構成

この節では、入力表と出力表の構成について説明します。

Text Search Plug-in Index Generator は、HiRDB に登録されている電子ドキュメントを取り出し、テキストデータを抽出します。この電子ドキュメントが登録されている表を入力表といいます。抽出したテキストデータは、全文検索用データとして HiRDB に登録されます。この全文検索用データが格納される表を出力表といいます。これら入力表および出力表の構成は、HiRDB サーバ構成に従う必要があります。Text Search Plug-in Index Generator を適用した HiRDB サーバ環境を新規に構築する場合は、入力表および出力表がその環境に適した表構成となるように設計してください。既存の HiRDB サーバ環境に Text Search Plug-in Index Generator を適用する場合は、入力表および出力表がその環境に適合するかどうかを検討してください。なお、全文検索用列が入力表にある場合と、異なる表にある場合では、表構成が異なります。

全文検索用列が入力表にある場合と異なる表にある場合での、入力表と出力表の構成を、表 2-1 ~ 2-3 に示します。

表 2-1 全文検索用列が入力表と同じである場合の表構成

項番	列	データ型 (データ長)	ナル属性	入出力	情報設定者	インデクス作成	列追加
1	ユニークキー列	ユーザ指定	不可	入力	ユーザ	推奨	既存列流用可
2	入力ドキュメント格納列	BINARY または BLOB	可	入力	ユーザ	不要	既存列流用可
3	入力ドキュメントファイル名称格納列	MVARCHAR (512 バイト)	可	入力	ユーザ	不要	既存列流用可
4	入力ドキュメントコンテンツタイプ情報列	VARCHAR (255 バイトまたは 256 バイト)	可	入力	ユーザ	不要	既存列流用可
5	インデクス作成指定情報列	INTEGER	不可	入力	ユーザ	推奨	既存列流用可
6	インデクス作成状態列	INTEGER	不可	入出力	初期化：ユーザ更新：TSPIndexer コマンド	推奨	追加要
7	TSPIndexer 保守情報列	VARCHAR (8 バイト)	不可	入出力	初期化：ユーザ更新：TSPIndexer コマンド	不要	追加要
8	全文検索用列	SGMLTEXT または FREEWORD	可	出力	TSPIndexer コマンド	必須	追加要

表 2-2 全文検索用列が入力表と異なる場合の入力表構成

項番	列	データ型 (データ長)	ナル属性	入出力	情報設定者	インデクス作成	列追加
1	ユニークキー列	ユーザ指定	不可	入力	ユーザ	推奨	既存列 流用可
2	入力ドキュメント格納列	BINARY または BLOB	可	入力	ユーザ	不要	既存列 流用可
3	入力ドキュメントファイル名称格納列	MVARCHAR (512 バイト)	可	入力	ユーザ	不要	既存列 流用可
4	入力ドキュメントコンテンツタイプ情報列	VARCHAR (255 バイトまたは 256 バイト)	可	入力	ユーザ	不要	既存列 流用可
5	インデクス作成指定情報列	INTEGER	不可	入力	ユーザ	推奨	既存列 流用可
6	インデクス作成状態列	INTEGER	不可	入出力	初期化：ユーザ 更新：TSPIndexer コマンド	推奨	追加要
7	TSPIndexer 保守情報列	VARCHAR (8 バイト)	不可	入出力	初期化：ユーザ 更新：TSPIndexer コマンド	不要	追加要

表 2-3 全文検索用列が入力情報表と異なる場合の出力表の構成

項番	列	データ型	ナル属性	入出力	情報設定者	インデクス作成	列追加
1	ユニークキー列	ユーザ指定	不可	入出力	ユーザ	推奨	既存列 流用可
2	全文検索用列	SGMLTEXT または FREEWORD	可	出力	TSPIndexer コマンド	必須	追加要

以降、表 2-1 から表 2-3 に示す列の詳細について説明します。

### (1) ユニークキー列

TSPIndexer コマンドがデータを操作するときに必要なユニークキーを指定します。各行で重複しないキー値を指定してください。ユニークキーを構成する列は、複数列でユニークとなる列群にも対応します。指定できる列数は3列までです。既存列を指定することもできます。指定できる列の属性を次に示します。

- SMALLINT
- INTEGER

## 2. 環境設定

- CHAR
- VARCHAR
- NCHAR
- NVARCHAR
- MCHAR
- MVARCHAR

### 注意事項

- 文字データ型と混在文字データ型列の長さは、最大 255 バイトです。
- 各国文字データ型の長さは、最大 127 文字です。
- 全文検索用列が入力情報表と異なる場合、出力表は入力表と同じ構成にしてください。入力表と同一のキーが出力表にあった場合は更新され、なかった場合は追加されます。
- TSPIndexer コマンドでは、値をチェックしませんが、値がユニークでない場合は結果を保証しません。
- ユニークキーに文字データを使用する場合は、半角文字コード、全角文字コード、TAB(0x09)、NL(0x0a)、CR(0x0d) を使用してください。これら以外のコードを使用した場合は、正常に処理されない場合があります。

### (2) 入力ドキュメント格納列

全文検索対象とする電子ドキュメントが格納されている列です。既存列を指定できます。指定できる列の属性を次に示します。

- BLOB
- BINARY

ナル値データまたはデータ長が 0 バイトのデータに対して、全文検索列への登録処理が実行された場合、行データ不正として当該行の処理をスキップします。

### (3) 入力ドキュメントファイル名称格納列

全文検索対象とする電子ドキュメントのファイル名を格納する列です。「ファイル名. 拡張子」の形式で格納してください。ファイル名が 512 バイトを超える場合は、ファイル名の先頭文字を削除して格納します。終端の拡張子は削除しません。ファイル名には . (ピリオド) を指定できます。

ナル値データまたはデータ長が 0 バイトのデータに対して、全文検索列への登録処理が実行された場合、行データ不正として当該行の処理をスキップします。

### (4) 入力ドキュメントコンテンツタイプ格納列

全文検索対象とする電子ドキュメントのコンテンツタイプを格納する列です。

### (5) インデクス作成指定情報列

全文検索用列への登録を実行するかどうか (全文検索対象とするかどうか) を指定しま



す。次のどちらかの値を指定してください。

0

全文検索用列への登録を実行しません。

1

全文検索用列への登録を実行します。

## (6) インデクス作成状態列

(5) でインデクス作成を指定された電子ドキュメントに対して、実際に登録を完了したかどうかを設定します。この列は、電子ドキュメントを登録または更新したときに、初期値で初期化されます。TSPIndexer コマンドは、初期状態の場合に登録処理を実行しません。

インデクス作成状態列には、次のどれかが設定されます。

設定値	内容
0	全文検索列へデータ未登録（初期値）
1	全文検索列へデータ登録済み
100	抽出したテキスト長が Text Search Plug-in の制限を超えました
101	入力行のデータ不正（サポート外文書含む）によって、登録に失敗しました
102	データ不正以外の要因によって、登録に失敗しました

抽出したテキストデータは、全文検索用列に登録されます。このとき、登録先の全文検索用列のデータ型種別によって、登録できるテキストサイズが次のように異なります。

- 登録先が SGMLTEXT 型列の場合：5,242,880 バイト
- 登録先が FREEWORD 型列の場合：32,000 バイト

抽出したテキストデータがこのサイズを超える場合は、LEVEL\_FORCE に指定した内容に従って処理されます。LEVEL\_FORCE の指定値とそれに対応する処理を次に示します。

### 0 を指定した場合

エラーとなります。該当行に対する登録処理をスキップし、インデクス作成状態列には「100（抽出したテキスト長が Text Search Plug-in の制限を超えました）」が設定されます。

### 1 を指定した場合

テキストの先頭から、格納可能な長さだけ格納します。インデクス作成状態列には「1（全文検索列へデータ登録済み）」が設定されます。

## (7) TSPIndexer 保守情報列

TSPIndexer コマンドの保守用の列です。電子ドキュメントの登録・更新時には、初期値

## 2. 環境設定

で初期化してください。それ以外の場合は更新しないでください。

### (8) 全文検索用列

電子ドキュメントから抽出したテキストデータを登録する、全文検索用列です。データ型は、SGMLTEXT 型または FREEWORD 型のどちらかを設定できます。NGRAM インデクスまたは IXFREEWORD インデクスは、Text Search Plug-in の遅延更新機能をご使用になることをお勧めします。全文検索用列のインデクス定義については、「2.3 インデクスの作成」を参照してください。

## 2.3 インデクスの作成

---

この節では、インデクス作成時に考慮していただきたいことについて説明します。

次の列は、インデクスを作成することで処理性能の向上を実現できます。

### ユニークキー列

#### インデクス作成状態列およびインデクス作成指定情報列

通常運用の場合は、インデクス作成状態列に対する単一列インデクスとしてください。インデクス作成指定情報列には 1 を設定します。

それ以外の場合は、インデクス作成状態列およびインデクス作成指定情報列を構成列とする複数列インデクスを定義する必要があります。

#### 注意

- インデクス作成状態列とインデクス作成指定情報列のうち、より絞り込める列を複数列インデクスの第一構成列としてインデクスを定義してください。

### 全文検索用列

インデクス作成は必須です。列の属性によって、定義するインデクスが異なります。

- SGMLTEXT 型列の場合  
NGRAM インデクスを定義します。
- FREEWORD 型列の場合  
IXFREEWORD インデクスを定義します。

全文検索用インデクスは、Text Search Plug-in のインデクス分割遅延登録機能を使用して、インデクス作成の遅延運用をお勧めします。Text Search Plug-in のインデクス分割遅延登録機能を利用した運用には、次のものがあります。

#### インデクス作成方法に初期分割登録を指定してインデクスを作成する運用

大量の文書または文字列データを表に登録したあと、空のインデクスだけを作成し、インデクス作成開始ユティリティを実行してインデクスへ文書または文字列データを登録する運用

#### インデクス作成方法に分割遅延登録を指定してインデクスを作成する運用

インデクスを作成しないで UAP でデータを登録したあと、一括してインデクスを作成する方法です。

「プラグインインデクスの遅延一括作成による運用」に比べて、次のような特長があります。

ユティリティを使用してインデクスを作成する場合、作成時間または作成件数を指定できるため、柔軟に運用できます。

遅延登録するかどうかを、列（インデクス）単位で指定できます。

登録・検索同時実行を併用することで、ユティリティによるインデクス作成中に検索が実行できます。

Text Search Plug-in のインデクス分割遅延登録機能の詳細については、マニュアル「HiRDB Text Search Plug-in Version 7」を参照してください。



# 3

## 全文検索用データの登録

この章では、TSPIndexer コマンドの機能について説明します。また、TSPIndexer コマンドの処理を定義する制御情報ファイルの定義方法についても説明します。

---

3.1 TSPIndexer (全文検索用データ登録コマンド)

---

3.2 制御情報ファイルの定義

---

## 3.1 TSPIndexer (全文検索用データ登録コマンド)

Text Search Plug-in Index Generator で電子ドキュメントから全文検索用データを抽出し、データベースに登録するには、TSPIndexer コマンド (全文検索用データ登録コマンド) を利用します。この節では、TSPIndexer コマンドの機能について説明します。

### (1) 機能

ワードプロセッサなどの文書作成ソフトウェアで作成した電子ドキュメントから、全文検索用テキストデータを抽出し、データベースに登録します。

### (2) 形式

TSPIndexer -f 制御情報ファイル -c 使用する文字コード

### (3) 引数

-f 制御情報ファイル

制御情報ファイル名をドライブ名からの絶対パスで指定します。制御情報ファイルとは、このコマンドの処理内容を定義したファイルのことです。制御情報ファイルの詳細については、「3.2 制御情報ファイルの定義」を参照してください。

-c 使用する文字コード

HIRDB システムおよび制御情報ファイルで使用する文字コードを指定します。文字コードは、次のどちらかを指定してください。

sjis : Sjis を使用する場合に指定します。

utf-8 : UTF-8 を使用する場合に指定します。

### (4) 戻り値

戻り値を次に示します。

戻り値	内容
0	正常終了 (登録に失敗した文書はありません)
1	登録に失敗した文書があります
-1	処理を続行できないエラーが発生しました

### (5) 注意事項

- TSPIndexer コマンドの実行者は、入力表および出力表に対してアクセス権が必要です。
- このコマンドを実行中に表の行が削除された場合、エラーメッセージを出力して処理を続行します。
- 表に排他を設定すると、トランザクションが終了するまでコマンドは実行待ちとなります。

ます。このため、表には排他を設定しないようにしてください。

## 3.2 制御情報ファイルの定義

この節では、TSPIndexer コマンドの処理内容を定義する制御情報ファイルの記述形式、定義方法について説明します。

### 3.2.1 制御情報ファイルの記述形式と規則

制御情報ファイルに定義する項目の記述形式と規則を次に示します。

#### 記述形式

設定項目=設定値 [ , 設定値 ] ...

#### 記述規則

次の規則に従って記述してください。規則に従っていない場合は、コマンドが正常に実行されません。

- コメントを付ける場合は、行頭に # を付けてください。
- 改行すると、その設定項目の指定が完了したとみなされます。
- 設定項目と設定値は半角の = (イコール) で結んでください。設定項目と設定項目、設定値と設定値の間には、空白を指定できません。
- 行頭から記述してください。
- 行間の区切り文字には、<CR>+<LF> または <LF> を使用してください。
- 英字の大文字と小文字は区別されます。
- 複数の設定値を指定する場合は、設定値と設定値を、(コンマ) で区切ってください。
- 列名または表名に半角英小文字を使用する場合は、" (ダブルクォーテーション) で囲んでください。TSPIndexer コマンドは、指定された名称でデータベースにアクセスするため、大文字・小文字を区別するために引用符が必要です。

### 3.2.2 制御情報ファイルの指定項目と指定方法

制御情報ファイルには、TSPIndexer コマンドの処理内容を定義します。制御情報ファイルはサンプルが提供されていますので、ご使用方法に合わせてカスタマイズしてください。

#### (1) 制御情報ファイルの指定項目

制御情報ファイルに指定する項目を次の表に示します。

表 3-1 制御情報ファイルに指定する項目

項番	指定項目	指定内容	指定種別	複数指定	指定要否
1	TBL_IN	入力表名	表名	不可	必須
2	TBL_OUT	出力表名	表名	不可	必須



項番	指定項目	指定内容	指定種別	複数指定	指定要否
3	COL_INDOC_UNIQUE	入力表ルのユニークキー列名および属性	列	可	必須
4	COL_INDOC_DOC	入力ドキュメントの格納列名	列	不可	必須
5	COL_INDOC_FNAME	入力ドキュメントファイル名の格納列名	列	不可	必須
6	COL_INDOC_CTYPE	入力ドキュメントのコンテンツタイプ格納列	列	不可	必須
7	COL_INDOC_MKSEARCH	インデクス作成指定情報列名	列	不可	必須
8	COL_MKSTAT	インデクス作成状態列名	列	不可	必須
9	COL_MNT	TSPIndexer 保守情報列名	列	不可	必須
10	COL_OUT_UNIQUE	出力表のユニークキー列名および属性	列	不可	必須
11	COL_OUT_TEXT	全文検索用列名	列	不可	必須
12	LEVEL_FORCE	登録指定以外の文書，または登録に失敗した文書を登録するときに指定する	数値	不可	必須
13	FILE_ERROR	エラー情報ファイル	ファイル名	不可	必須
14	TIME_EXEC	実行時間指定	数値	不可	任意
15	HIRDB_VERSION	使用する HiRDB のバージョン	文字列	不可	必須
16	FILE_DOCFILTER_CONF	Document Filter for Text Search のコンフィグレーションファイル名	ファイル名	不可	任意

制御情報ファイルに指定する項目は、指定値の種別によって指定方法が異なります。次に、指定値の種別ごとに指定方法を説明します。

(a) 表を指定する場合

入力表名を指定します。表名は、「認可識別子.表識別子」の形式で指定します。認可識別子は省略できません。

指定例

TBL\_IN=USER01.文書表

### 3. 全文検索用データの登録

TBL\_OUT="user02"."SearchTable"

#### (b) 列を指定する場合

列名, 列属性, 列長, および非ナル値制約オプションを指定します。指定する列の詳細については、「2.2 入力表および出力表の構成」を参照してください。なお, 指定した値と実際の表の形式が不一致の場合はエラーとなります。また, 代入, 比較などの変換できるデータ型同士の場合でもエラーとなりますのでご注意ください。

#### 指定形式

設定項目=列名, 列属性[, [列長] [, notnull]]

列を指定する場合の注意事項を次に示します。

- 複数の列を指定する場合は, 設定項目を複数指定してください。
- 列長の単位は, 各国文字データを使用している場合は文字数で指定してください。各国文字データ以外の場合は, バイト数で指定してください。
- INTEGER などの列長を必要としない列属性には, 列長の指定は不要です。このような場合に列長を指定すると, エラーとなります。
- 列にナル値の登録を許可しない場合は, 末尾に "notnull" を追加してください。  
"notnull" を指定しないと, ナル値の登録が許可されます。ただし, ナル属性不可で固定である列に対しては, "notnull" の指定は不要です。このような列に "notnull" を指定した場合はエラーとなります。
- 列属性は, 次の中から指定してください。

列属性	列長要否
SMALLINT	不要
INTEGER	不要
CHAR	必須
VARCHAR	必須
NCHAR	必須
NVARCHAR	必須
MCHAR	必須
MVARCHAR	必須
BLOB	必須
BINARY	必須
SGMLTEXT	不要
FREWORD	不要

- 設定項目は必ず形式に従って指定してください。途中の項目を省略する場合でも, 前後の, (コンマ) は必要です。任意の位置から以降すべてを省略する場合は, 指定項目以降の, (コンマ) の指定は不要です。

#### 指定例

```
COL_INDOC_UNIQUE=NUMBER, INTEGER
COL_INDOC_UNIQUE="name", CHAR, 128
COL_INDOC_DOC=DOC, BLOB, 1048576, notnull
COL_OUT_TEXT=TEXT_OUT, SGMLTEXT, , notnull
```

#### (c) ファイル名を指定する場合

ファイル名をドライブ名からの絶対パスで指定します。ファイル名の前後を" (ダブルクォーテーション) で囲んでください。

#### 指定例

```
FILE_ERROR="D:¥tmp¥error.log"
FILE_TRACE="D:¥trace¥trace.log"
```

### (2) 指定項目の詳細

制御情報ファイルの各設定項目について説明します。

#### (a) TBL\_IN

入力情報格納列が存在する表名を指定します。指定方法については、「3.2.2(1)(a) 表を指定する場合」を参照してください。

#### (b) TBL\_OUT

抽出したテキストデータを格納する全文検索用列の存在する表名を指定します。指定方法については、「3.2.2(1)(a) 表を指定する場合」を参照してください。

出力表と入力表が同じ場合は、「TBL\_IN」に指定した表名を指定してください。

#### (c) COL\_INDOC\_UNIQUE

入力表のユニークキー列名および属性を指定します。ユニークキー列は三つまで指定できます。三つを超えて指定した場合はエラーとなります。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。

ユニークキー列を複数指定する場合は、この設定項目を複数行指定してください。複数のユニークキー列を指定した場合、ユニークキーの優先順位は指定順となります。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (d) COL\_INDOC\_DOC

全文検索用テキストデータを抽出する電子ドキュメントの実体が格納されている列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (e) COL\_INDOC\_FNAME

全文検索用テキストデータを抽出する電子ドキュメントのファイル名が格納されている列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

### 3. 全文検索用データの登録

い。

#### (f) COL\_INDOC\_CTYPE

全文検索用テキストデータを抽出する電子ドキュメントのコンテンツタイプが格納されている列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (g) COL\_INDOC\_MKSEARCH

インデクス作成指定情報列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (h) COL\_MKSTAT

インデクス作成状態列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (i) COL\_MNT

TSPIndexer 保守情報列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (j) COL\_OUT\_UNIQUE

入力表のユニークキー列名および属性を指定します。ユニークキー列は三つまで指定できます。三つを超えて指定した場合はエラーとなります。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。

ユニークキー列を複数指定する場合は、この設定項目を複数行指定してください。複数のユニークキー列を指定した場合、ユニークキーの優先順位は指定順となります。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

各列の属性および指定順序は、COL\_INDOC\_UNIQUE と同じにしてください。

#### (k) COL\_OUT\_TEXT

全文検索列名を指定します。指定方法については、「3.2.2(1)(b) 列を指定する場合」を参照してください。列の詳細については、「2.2 入力表および出力表の構成」を参照してください。

#### (l) LEVEL\_FORCE

テキスト抽出・全文検索列登録する行の実行レベルを指定します。次のどれかを指定してください。

0

インデクス作成指定情報列で「1 (登録を実行する)」が指定され、さらに、インデクス作成状態列が0 (未登録) である行を処理対象とします。

1

インデクス作成指定情報列で「1 (登録を実行する)」が指定され、さらに、インデクス作成状態列が0 (未登録) である行を処理対象とします。また、抽出したテキスト長がプラグインの制限を超える場合、テキストの先頭から格納できる長さだけを格納します。このとき、以前にテキスト長がプラグインの制限を超えてエラーとなった行 (インデクス作成状態が100) も登録します。

2

インデクス作成指定情報列で「1 (登録を実行する)」が指定され、さらに、インデクス作成状態列が0 (未登録) である行を処理対象とします。また、インデクス状態列が102 (データ不正以外の要因で登録を失敗した) の行を処理対象とします。データ不正以外の障害が発生した場合の、再実行用として利用します。

3

1 および 2 の機能を含みます。

#### (m) FILE\_ERROR

障害発生時のエラー情報および TSPIndexer コマンドの実行開始・実行完了メッセージを出力する、出力先ファイル名を指定します。ファイル名はドライブ名からの絶対パスで指定してください。なお、ファイルの出力先となるディレクトリは事前に作成しておく必要があります。指定方法については、「3.2.2(1)(c) ファイル名を指定する場合」を参照してください。

TSPIndexer コマンドを実行すると、指定されたファイルに情報が追加されます。

#### (n) TIME\_EXEC

TSPIndexer コマンドの実行時間を、0 ~ 14400 の分単位で指定します。デフォルトは60です。TSPIndexer コマンドは、ここで指定された時間だけ、登録処理を実行します。0を指定すると無制限と解釈され、すべての登録対象列に対して処理を実行します。

#### (o) HIRDB\_VERSION

TSPIndexer コマンドで使用する、HiRDB のバージョンを指定します。6文字の半角英数字で指定してください。-(ハイフン) や/(スラッシュ) は指定できません。

また、バージョンアップなどで HiRDB のバージョンが変更になった場合は、この指定値も変更してください。

#### 指定例

06-01 の場合：060100

07-00-/A の場合：07000A

### 3. 全文検索用データの登録

07-00-2C の場合 : 07002C

06-01-A の場合 : 06010A

#### (p) FILE\_DOCFILTER\_CONF

Document Filter for Text Search の動作を定義したコンフィグレーションファイル名をドライブ名からの絶対パスで指定します。TSPIndexer コマンドがテキストデータを抽出するときに、Document Filter for Text Search はここで指定したコンフィグレーションファイルの内容に従って動作します。指定しなかった場合は、Document Filter for Text Search での定義内容に従って動作します。また、コンフィグレーションファイルに指定されていない項目も、Document Filter for Text Search での定義内容に従います。

なお、指定したコンフィグレーションファイルを READ モードで OPEN できなかった場合は、エラーメッセージ KFGX2001-E を出力して処理を中断します。

コンフィグレーションファイル ( config.cfg ) の設定例を次に示します。

```
[ SYSTEM ]
  DOCFTIMEOUT      =600

[ DOCFLIBRARYMAP ]
  csv               =DOCF
  ini               =DOCF
  cfg               =DOCF
  txt               =DOCF
  csv               =DOCF

[ DOCFTIPEMAP ]
  csv               =csv
  ini               =ini
  cfg               =cfg
  .                 =txt
```

### 3.2.3 制御情報ファイルの定義例

制御情報ファイルの定義例を次に示します。

```
#####  
##                               ##  
## sample_control.txt           ##  
##                               ##  
## Control information file for HiRDB Text Search Plug-in Index Generator ##  
##                               ##  
## Copyright (C) 2004, Hitachi, Ltd.           ##  
##                               ##  
#####  
TBL_IN="root".TBL_IN  
TBL_OUT="root".TBL_OUT  
COL_INDOC_UNIQUE=KEY_IN,INTEGER  
COL_INDOC_DOC=DOC,BLOB,1048576  
COL_INDOC_FNAME=FNAME,MVARCHAR,512  
COL_INDOC_CTYPE=CTYPE,VARCHAR,256  
COL_INDOC_MKSEARCH=MKSEARCH,INTEGER  
COL_MKSTAT=MKSTAT,INTEGER  
COL_MNT=MNT,VARCHAR,8  
COL_OUT_UNIQUE=KEY_OUT,INTEGER  
COL_OUT_TEXT=TEXT_OUT,SGMLTEXT  
LEVEL_FORCE=0  
FILE_ERROR="c:\¥error.txt"  
#TIME_EXEC=60  
HIRDB_VERSION=06010C  
#FILE_DOCFILTER_CONF="c:\¥config.cfg"
```





# 4

## 運用

この章では、Text Search Plug-in Index Generator の効率的な運用方法について説明します。

---

4.1 TSPIndexer コマンド実行時の留意点

---

4.2 推奨する運用方法

---

## 4.1 TSPIndexer コマンド実行時の留意点

---

TSPIndexer コマンドを効率良く実行するための留意点を次に説明します。

### (1) 排他待ち時間の確保

ある行に対し、更新トランザクション処理と TSPIndexer コマンドの処理が同時となった場合、あとから実行されたトランザクションが排他待ちとなります。このため、排他待ち時間は十分に確保する必要があります。

### (2) TSPIndexer コマンドの同時実行について

一つの表に対しては、一つの TSPIndexer コマンドが処理対象とするようにしてください。一つの表に対して同時に複数の TSPIndexer コマンドを実行した場合、TSPIndexer コマンドごとに対象部分に排他が設定されます。このため、性能が低下したり、排他時間が長くなるためにほかの運用に影響を及ぼしたりする可能性があります。

TSPIndexer コマンドが同時に実行され、各コマンドが同じ行を更新する場合には、処理対象一覧は重複して作成されますが、各行を更新する前に、再度更新可否のチェックを実施します。このため、最初の更新要求を処理し、それ以降の更新要求はスキップするので、データが二重に登録されることはありません。

### (3) デッドロックについて

TSPIndexer コマンドは、HiRDB のアプリケーションプログラムとして動作します。このため、同時に実行される他のトランザクションに排他状態によって、デッドロックが発生する可能性があります。

入力表と出力表が異なる運用では、TSPIndexer コマンドは次の順序で発行されます。

1. 入力対象行を Exclusive Lock で排他
2. 出力表の該当行に対する UPDATE による排他

この順序の逆にほかのトランザクションから排他が行われる場合に、デッドロックが発生する可能性があります。

## 4.2 推奨する運用方法

Text Search Plug-in で適切な設定をすることで、Text Search Plug-in Index Generator を利用したシステムの登録運用を効率良く実行できます。

この節では、Text Search Plug-in での設定内容と、どのような場合に適用すれば効率がよくなるのかを説明します。なお、Text Search Plug-in での設定内容の詳細については、マニュアル「HiRDB Text Search Plug-in Version 2」または「HiRDB Text Search Plug-in Version 7」を参照してください。

Text Search Plug-in で推奨するパラメタの組み合わせを次の表に示します。

表 4-1 Text Search Plug-in で推奨するパラメタの組み合わせ

項番	create index 文のプラグインオプション		環境定義ファイルの設定項目		phnstartidx, phxstartidx の -w オプション
			phs_search_wait, phc_search_wait	phn_increment_practical_use, phx_increment_practical_use	
1	CREATEMODE=2	FLEX_AR EA	wait	off	wait
2	CREATEMODE=2	FLEX_AR EA	nowait	off	nowait
3	CREATEMODE=3	FLEX_AR EA	nowait_with_update	off	nowait_with_update

### 注

HiRDB Version 7 および HiRDB Text Search Plug-in Version 7 を使用している場合だけ設定できます。

#### 1 に適した業務

NGRAM 型のインデクス作成または更新をバッチで実行し、かつ NGRAM 型のインデクス作成中にほかの業務を実行しない業務に適しています。なお、この業務では、データ更新と同時に全文検索または概念検索を実行する必要がないことが条件となります。

#### 2 に適した業務

NGRAM 型のインデクス作成または更新をバッチで実行し、かつ NGRAM 型のインデクス作成中に検索を実行する業務に適しています。この場合、検索時には、SELECT 文の排他オプションには without\_lock\_nowait だけが指定できます。なお、この業務では、データ更新と同時に全文検索または概念検索を実行する必要がないことが条件となります。

## 4. 運用

### 3 に適した業務

データの更新から少しあと（10分～1時間後）に、全文検索または概念検索を実行する業務に適しています。また、NGRAM 型のインデクス作成、更新、または更新中に更新系 SQL を実行する必要がある業務にも適しています。

# 5

## 障害対策

この章では、TSPIndexer コマンド実行時に発生した障害への対処方法について説明します。また、エラーメッセージとその対処方法についても説明します。

---

5.1 コマンド実行時の障害対策

---

5.2 メッセージ

---

## 5.1 コマンド実行時の障害対策

---

この節では、TSPIndexer コマンドで障害が発生した場合の対処方法について説明します。

### 5.1.1 障害の切り分け方法

障害が発生した場合、次の観点で対処してください。

1. Text Search Plug-in Index Generator のエラーログと、HiRDB サーバの syslog ( イベントログ ) を比較し、最初に発生したエラー ( 発生時間の一番古いエラー ) を探してください。
2. 1. で探したエラーが Text Search Plug-in Index Generator のエラーの場合は、Text Search Plug-in Index Generator が出力したエラーメッセージに従って対処してください。
3. 1. で発生したエラーが HiRDB サーバの syslog ( イベントログ ) に出力されている場合は、そのエラーを発生させたプログラムのメッセージに従って対処してください。

### 5.1.2 障害発生時の対策方法

#### (1) HiRDB で障害が発生した場合

HiRDB で障害が発生した場合、障害が発生した行以降は未登録の状態となります。TSPIndexer コマンドが出力したエラー情報ファイル、または HiRDB サーバのエラー情報を確認してください。障害要因を取り除いたあと、再度 TSPIndexer コマンドを実行してください。

#### (2) 入力データが不正であった場合

入力データが Document Filter for Text Search が処理できない不正データやサポート外文書であった場合、および入力指定された各種の列情報が不正であった場合は、エラーとなります。この場合、エラー情報ファイルにエラー要因とユニークキー列の情報が出力され、インデクス作成状態列に 101 が設定されます。ユニークキーからエラーとなった行を特定できますので、該当する行を削除するか、更新してから、再度 TSPIndexer コマンドを実行してください。行を更新しない場合は、TSPIndexer コマンドの処理対象となりません。

#### (3) Document Filter for Text Search の環境不正によってエラーが発生した場合

メモリ不足などの Document Filter for Text Search の環境不正が要因でエラーが発生した場合、TSPIndexer コマンドは登録に失敗した行をスキップし、次の行から処理を続行します。このとき、インデクス作成状態列に 102 が設定されます。

エラーメッセージを確認して、障害要因を削除してください。そのあと、LEVEL\_FORCE に 2 を指定して、TSPIndexer コマンドを再実行してください。

#### (4) TSPIndexer コマンドを続行できないエラーが発生した場合

定義ファイルが不正の場合や、Text Search Plug-in Index Generator のメモリ不足の場合、TSPIndexer コマンドは処理を続行できません。エラーメッセージファイルおよび標準出力に出力されたエラー情報を確認し、障害要因を削除してください。そのあと、TSPIndexer コマンドを再実行してください。

## 5.2 メッセージ

---

この節では、TSPIndexer コマンドが出力するメッセージの形式、および出力内容について説明します。

### 5.2.1 メッセージの形式と見方

#### (1) メッセージの出力先

パラメタ不正や制御情報ファイル不正の場合は、画面にメッセージを出力します。登録処理でエラーが発生した場合は、エラー情報ファイルにメッセージを出力します。エラー情報ファイルの出力先については、「3.2.2(2)(m) FILE\_ERROR」を参照してください。

#### (2) メッセージの形式

TSPIndexer コマンドが出力するメッセージの形式を次に示します。

KFGXnnnn-X：メッセージテキスト

KFGX

メッセージを出力したプログラム (Text Search Plug-in Index Generator) を表します。

nnnn

メッセージの番号です。それぞれのメッセージに 4 けたの固有の番号が付いています。

X

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、処理を中止します。

#### (3) メッセージの見方

このマニュアルでは、メッセージを次に示す形式で説明しています。

#### **KFGXnnnn-X メッセージテキスト**

---

メッセージの意味を説明しています。

[ 処理 ] Text Search Plug-in Index Generator の処理を示します。

[ 対策 ] メッセージを受け取った HiRDB 管理者の処置を示します。

なお、[ 対策 ] の「保守員に連絡してください」とは、購入時の契約に基づいて、システム管理者が弊社問い合わせ窓口へ連絡することを示します。



## 5.2.2 メッセージ一覧

### **KFGX0001-I Usage: TSPIndexer -f Control\_File -c {sjis|utf-8}**

---

TSPIndexer コマンドの形式を示します。

[ 対策 ]

TSPIndexer コマンドの形式に従って実行してください。

### **KFGX0002-I TSPIndexer started. aa..aa**

---

TSPIndexer コマンドの実行が開始されました。

aa..aa : コマンド開始日時

[ 対策 ]

なし

### **KFGX0003-I TSPIndexer ended. rc=XX aa..aa**

---

TSPIndexer コマンドが終了しました。

XX : リターンコード

aa..aa : コマンド終了日時

[ 対策 ]

なし

### **KFGX0004-I Timeout occurred. Processed number of documents : XX**

---

タイムアウトが発生しました。

XX : 処理した件数

[ 対策 ]

コマンドを再実行してください。

### **KFGX0005-I Error occurred. Processed number of documents : XX**

---

エラーが発生しました。

XX : 処理した件数

[ 対策 ]

エラー情報ファイルに出力されているエラーメッセージの対策方法に従って対処してください。

### **KFGX0006-I All the processing completed. Processed number of documents : XX**

---

すべての処理を完了しました。

XX : 処理した件数

[ 対策 ]

なし

### **KFGX0007-I Processed number of documents : XX**

---

処理した件数を示します。

XX : 処理した件数

[ 対策 ]

なし

### **KFGX2000-E Control file name is too long.**

---

制御情報ファイル名が長過ぎます。

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイル名を短く指定し直してから、再実行してください。

### **KFGX2001-E File open error. file=aa..aa errno=XX**

---

ファイルのオープンに失敗しました。

aa..aa : ファイルパス

XX : エラー番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

オープンに失敗したファイルを見直してください。

### **KFGX2002-E Syntax error. line=XX**

---

シンタックスエラーが発生しました。

XX : エラーの発生した制御情報ファイルの行番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

### **KFGX2003-E Definitions is too many.(aa..aa)**

---

同じ設定項目を、制限数を超えて定義しています。

aa..aa : 設定項目名

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2004-E Invalid item. line=XX**

---

設定項目名が不正です。

XX：エラーの発生した制御情報ファイルの行番号

## [ 処理 ]

TSPIndexer コマンドの実行を中断します。

## [ 対策 ]

制御情報ファイルを修正してください。

**KFGX2005-E aa..aa is not defined.**

---

aa..aa が定義されていません。

aa..aa：設定項目名

## [ 処理 ]

TSPIndexer コマンドの実行を中断します。

## [ 対策 ]

制御情報ファイルを修正してください。

**KFGX2006-E COL\_INDOC\_UNIQUE,COL\_OUT\_UNIQUE unmatched.**

---

COL\_INDOC\_UNIQUE と COL\_OUT\_UNIQUE の指定数，属性，または列長が不一致です。

## [ 処理 ]

TSPIndexer コマンドの実行を中断します。

## [ 対策 ]

制御情報ファイルを修正してください。

**KFGX2007-E Invalid value.(aa..aa) line=XX offset=YY**

---

設定値が不正です。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

YY：エラーの発生した行のオフセット位置

## [ 処理 ]

TSPIndexer コマンドの実行を中断します。

## [ 対策 ]

制御情報ファイルを修正してください。

**KFGX2008-E Column name is not defined.(aa..aa) line=XX**

---

列名指定がありません。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

## [ 処理 ]

## 5. 障害対策

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

## **KFGX2009-E Attribute is not defined.(aa..aa) line=XX**

---

属性指定がありません。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

## **KFGX2010-E Column name is too long.(aa..aa) line=XX**

---

列名指定が長過ぎます。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

## **KFGX2011-E Invalid attribute.(aa..aa) line=XX**

---

属性指定が不正です。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

## **KFGX2012-E Column length is not defined.(aa..aa) line=XX**

---

列長指定がありません。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

**KFGX2013-E Invalid number.(aa..aa) line=XX**

---

数値指定が不正です。

aa..aa : 設定項目名

XX : エラーの発生した制御情報ファイルの行番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2014-E File name is too long.(aa..aa) line=XX**

---

ファイル名指定が長過ぎます。

aa..aa : 設定項目名

XX : エラーの発生した制御情報ファイルの行番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2017-E Cannot define same column name.(aa..aa)**

---

ユニークキーに同じ列名は指定できません。

aa..aa : 設定項目名

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2018-E HiRDB (aa..aa) error occurred. message=(bb..bb)**

---

HiRDB でエラーが発生しました。

aa..aa : コマンド名

bb..bb : HiRDB のエラーメッセージ

[ 処理 ]

KFGX2020-E または KFGX2024-E が同時に出力されている場合は、処理失敗行に対する処理は中止し、TSPIndexer コマンドは処理を続行します。それ以外の場合は、処理を中断します。

[ 対策 ]

マニュアル「HiRDB メッセージ」を参照し、問題を取り除いてください。

**KFGX2019-E HiRDB SQLDA error occurred.**

---

HiRDB の SQLDA 作成部分でエラーが発生しました。

## 5. 障害対策

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

保守員に連絡してください。

## **KFGX2020-E Document type error.**

---

登録するドキュメントの種別が不正です。

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

保守員に連絡してください。

## **KFGX2021-E Invalid table name.(aa..aa) line=XX**

---

表名称指定が不正です。

aa..aa：設定項目名

XX：エラーの発生した制御情報ファイルの行番号

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを修正してください。

## **KFGX2023-E Column information error. table=aa..aa column=bb..bb**

---

制御情報ファイルに指定したカラム情報が不正です。

aa..aa：テーブル名

bb..bb：カラム名

### [ 処理 ]

TSPIndexer コマンドの実行を中断します。

### [ 対策 ]

制御情報ファイルを見直し、問題を取り除いてください。

## **KFGX2024-E Fail to INSERT/UPDATE line. Unique key information : aa..aa**

---

行のINSERT/UPDATEに失敗しました。

aa..aa：ユニークキー情報

### [ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行します。

### [ 対策 ]

問題を取り除いて、再度実行してください。

**KFGX2025-E Invalid version.(aa..aa) line=XX**

---

バージョン番号指定が不正です。

aa..aa : 設定項目名

XX : エラーの発生した制御情報ファイルの行番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2026-E Invalid file name.(aa..aa) line=XX**

---

ファイル名指定が不正です。

aa..aa : 設定項目名

XX : エラーの発生した制御情報ファイルの行番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

制御情報ファイルを修正してください。

**KFGX2027-E Invalid line information. column=aa..aa, Unique Key=information:bb..bb**

---

行の情報が不正です。

aa..aa : 不正を検出した行の列名

bb..bb : 不正を検出した行のユニークキー情報

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行します。

[ 対策 ]

不正を検出した行のデータを修正してください。

**KFGX2100-E DOCF\_Init error [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX : エラーコード

aa..aa : エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行します。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2101-E DOCF\_SetOption error ALL-OUTPUT [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2102-E DOCF\_SetOption error ALL-MODE [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2103-E DOCF\_OpenDocument error [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2104-E DOCF\_OpenContent error [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しま



す。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

---

**KFGX2105-E DOCF\_GetFirstText error [XX]: aa..aa**

Document Filter for Text Search でエラーが発生しました。

XX : エラーコード

aa..aa : エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

---

**KFGX2106-E DOCF\_GetNextText error [XX]: aa..aa**

Document Filter for Text Search でエラーが発生しました。

XX : エラーコード

aa..aa : エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

---

**KFGX2107-E DOCF\_CloseContent error [XX]: aa..aa**

Document Filter for Text Search でエラーが発生しました。

XX : エラーコード

aa..aa : エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2108-E DOCF\_CloseDocument error [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX2109-E DOCF\_Term error [XX]: aa..aa**

---

Document Filter for Text Search でエラーが発生しました。

XX：エラーコード

aa..aa：エラーメッセージ

[ 処理 ]

処理を失敗した行に対する処理を中止し、TSPIndexer コマンドは処理を続行しません。

[ 対策 ]

Document Filter for Text Search のドキュメントを参照してエラー要因を取り除いてください。

### **KFGX9000-E Not enough memory. size=XX**

---

メモリ不足が発生しました。

XX：確保しようとしたメモリサイズ (バイト)

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

システムのメモリ状態を確認と、再実行してください。

### **KFGX9001-E File I/O error. file=aa..aa func=bb..bb errno=XX**

---

ファイル I/O エラーが発生しました。

aa..aa：エラーが発生したファイルのパス

bb..bb：エラーが発生した関数

XX：エラー番号

[ 処理 ]

TSPIndexer コマンドの実行を中断します。

[ 対策 ]

ファイルのアクセス権やディスク容量などを確認してください。

# 付録

---

付録A ディレクトリ構成

---

付録B 用語解説

---

## 付録 A ディレクトリ構成

Text Search Plug-in Index Generator のディレクトリ構成を OS ごとに示します。

図 A-1 Text Search Plug-in Index Generator のディレクトリ構成 (Windows の場合)

(システムディレクトリ) ¥Program Files¥HITACHRI

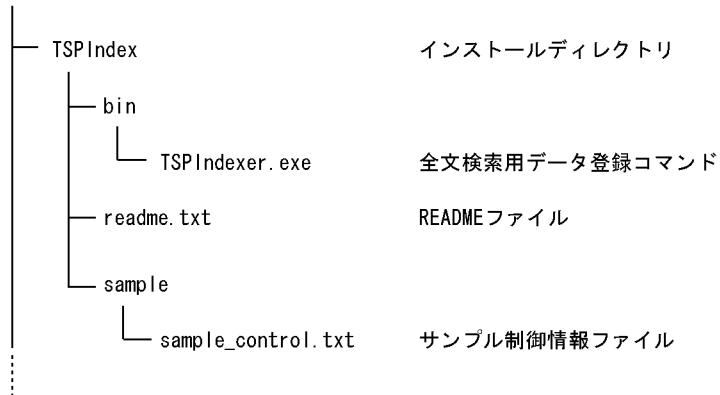
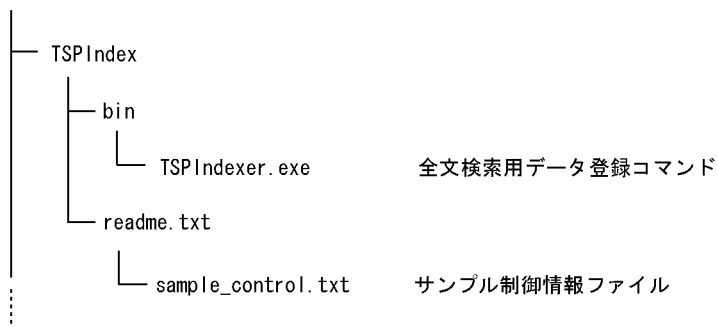


図 A-2 Text Search Plug-in Index Generator のディレクトリ構成 (UNIX の場合)

/opt



---

## 付録 B 用語解説

---

### (英字)

---

#### BLOB

文書や画像、音声など、キロバイト、メガバイト、またはギガバイト単位の長大なデータのことで  
す。

#### FREEWORD 型

文字列データを扱う抽象データ型のことです。

#### IXFREEWORD 型

FREEWORD 型という抽象データ型に対応するインデクス型のことです。

#### NGRAM 型

SGMLTEXT 型という抽象データ型に対応するインデクス型のことです。

#### SGMLTEXT 型

SGML 文書やプレーン文書を扱う抽象データ型のことです。

---

### (ア行)

---

#### インデクス

表を検索するためのキーとして列に付けた索引のことで、キーとキー値から構成されます。キーと  
は列の内容を示した列名のことで、キー値とは列の値のことで

インデクスには、単一列インデクスと複数列インデクスがあります。単一列インデクスとは、表の  
一つの列に作成した一つのインデクスのことです。また、複数列インデクスとは、表の複数の列で  
作成した一つのインデクスのことです。

---

### (カ行)

---

#### 更新前ログ取得モード

UAP またはユティリティを実行するときのデータベースの更新ログ取得方式の一つです。UAP ま  
たはユティリティが RD エリアの内容を更新するときに、ロールバックに必要なデータベース更新  
ログだけを取得する方式のことで

---

### (ナ行)

---

#### ナル値

値が設定されていないことを示す特殊な値のことで

## ( 八行 )

---

### 非ナル値制約

非ナル値制約とは、列の値にナル値を許さない制約のことです。

## ( ラ行 )

---

### ログ取得モード

UAPまたはユティリティを実行するときのデータベースの更新ログ取得方式の一つです。UAPまたはユティリティがRDエリアの内容を更新するときに、ロールバックおよびロールフォワードに必要なデータベース更新ログを取得する方式のことです。

### ログレスモード

UAPまたはユティリティを実行するときのデータベースの更新ログ取得方式の一つです。UAPまたはユティリティがRDエリアの内容を更新するときに、データベース更新ログを取得しない方式のことです。

---

# 索引

## B

---

BLOB 49

## C

---

COL\_INDOC\_CTYPE 24  
COL\_INDOC\_DOC 23  
COL\_INDOC\_FNAME 23  
COL\_INDOC\_MKSEARCH 24  
COL\_INDOC\_UNIQUE 23  
COL\_MKSTAT 24  
COL\_MNT 24  
COL\_OUT\_TEXT 24  
COL\_OUT\_UNIQUE 24

## F

---

FILE\_DOCFILTER\_CONF 26  
FILE\_ERROR 25  
FREEWORD 型 49

## H

---

HIRDB\_VERSION 25

## I

---

IXFREEWORD 型 49

## L

---

LEVEL\_FORCE 24

## N

---

NGRAM 型 49

## S

---

SGMLTEXT 型 49

## T

---

TBL\_IN 23

TBL\_OUT 23

Text Search Plug-in Index Generatorの機能  
2

Text Search Plug-in Index Generator のシス  
テム構成 4

Text Search Plug-in Index Generator の処理  
概要 3

Text Search Plug-in で推奨するパラメタの組  
み合わせ 31

TIME\_EXEC 25

TSPIndexer 18

TSPIndexer コマンド実行時の留意点 30

TSPIndexer 保守情報列 13

## い

---

インデクス 49

インデクス作成指定情報列 12

インデクス作成状態列 13

インデクスの作成 15

## か

---

環境変数の設定 8

## こ

---

更新前ログ取得モード 49

コマンド実行時の障害対策 34

## し

---

実行環境の設定 8

障害の切り分け方法 34

障害発生時の対策方法 34

## す

---

推奨する運用方法 31

## せ

---

制御情報ファイルに指定する項目 20

制御情報ファイルの記述形式と規則 20

制御情報ファイルの指定項目と指定方法 20  
制御情報ファイルの定義 20  
制御情報ファイルの定義例 26  
全文検索用データ登録コマンド 18  
全文検索用列 14

## て

---

ディレクトリ構成 48

## な

---

ナル値 49

## に

---

入カドキュメント格納列 12  
入カドキュメントコンテンツタイプ格納列  
12  
入カドキュメントファイル名称格納列 12  
入力表および出力表の構成 10

## ひ

---

非ナル値制約 50

## め

---

メッセージ 36  
メッセージ一覧 37  
メッセージの形式 36  
メッセージの形式と見方 36  
メッセージの出力先 36  
メッセージの見方 36

## ゆ

---

ユニークキー列 11

## ろ

---

ログ取得モード 50  
ログレスモード 50





(株) 日和 出版センター 行き

FAX 番号 0120-210-454 (フリーダイヤル)

## 日立マニュアル注文書

ご注文日	年 月 日
送付先ご住所	〒 ..... ..... .....
お客様名 (団体名, 又は法人名など)	
お名前	
電話番号	( )
FAX 番号	( )

資料番号	マニュアル名	数量
合計		

マニュアルのご注文について、ご不明な点は  
(株) 日和 出版センター (☎03-5281-5054) へお問い合わせください。