

HiRDB 全文検索プラグイン

HiRDB Text Search Plug-in Version 8

解説・手引書

3020-6-375-30

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

適用 OS : AIX 5L V5.1 , AIX 5L V5.2 , AIX 5L V5.3 , AIX V6.1

P-1MD3-5381 HiRDB Text Search Plug-in Version 8(64) 08-05

P-F1MD3-33711 HiRDB Text Search Plug-in Conceptual Extension Version 7 07-00

適用 OS : Red Hat Enterprise Linux AS 3 (AMD64 & Intel EM64T) , Red Hat Enterprise Linux AS 4(AMD64 & Intel EM64T) , Red Hat Enterprise Linux ES 4(AMD64 & Intel EM64T) , Red Hat Enterprise Linux 5 Advanced Platform(AMD/Intel 64) , Red Hat Enterprise Linux 5(AMD/Intel 64)

P-9WD3-5381 HiRDB Text Search Plug-in Version 8(64) 08-05

P-F9SD3-33711 HiRDB Text Search Plug-in Conceptual Extension Version 7 07-00

注 動作環境としては、Intel EM64T にだけ対応しています。

適用 OS : Windows 2000 , Windows XP Professional , Windows XP x64 Edition , Windows Server 2003 , Windows Server 2003 x64 Edition , Windows Vista Business , Windows Vista Enterprise , Windows Vista Ultimate , Windows Vista Business (x64) , Windows Vista Enterprise (x64) , Windows Vista Ultimate (x64) , Windows Server 2008 , Windows Server 2008 (x64)

P-24D3-3384 HiRDB Text Search Plug-in Version 8 08-05

P-24D3-9374 HiRDB Text Search Plug-in Conceptual Extension Version 7 07-00

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「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

AIX は、米国における米国 International Business Machines Corp. の登録商標です。

Itanium は、アメリカ合衆国および他の国におけるインテル コーポレーションまたはその子会社の登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標あるいは商標です。

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。

Red Hat は、米国およびその他の国で Red Hat, Inc. の登録商標若しくは商標です。

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 Microsoft Corporation の米国及びその他の国における登録商標です。

Windows Vista は、米国 Microsoft Corporation の米国及びその他の国における登録商標です。

発行

2007 年 8 月 (第 1 版) 3020-6-375

2009 年 1 月 (第 4 版) 3020-6-375-30

著作権

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

変更内容

変更内容 (3020-6-375-30) HiRDB Text Search Plug-in Version 8 08-05

| 追加・変更内容 | 変更箇所 |
|--|-------|
| メモリ所要量の見積もり式を変更しました。 | 2.3.1 |
| RD エリア容量の見積もり式を変更しました。 | 2.3.2 |
| IXFREEWORD インデクス型プラグインのインデクス定義のプラグインオプションに登録件数指定を追加しました。 | 2.7.4 |

単なる誤字・脱字などはお断りなく訂正しました。

変更内容 (3020-6-375-20) HiRDB Text Search Plug-in Version 8 08-04

| 追加・変更内容 |
|---|
| n-gram インデクス情報格納用 RD エリアの説明を変更しました。 |
| IXFREEWORD インデクス情報格納用 RD エリアの説明を変更しました。 |
| 注意事項を追加しました。 |
| 登録の環境を設定できる単位を変更しました。 |
| 注意事項を変更しました。 |
| かたかな異表記展開の結果の説明を変更しました。 |
| ディレクトリおよびファイル構成を変更しました。 |
| インデクス定義ファイルの説明を変更しました。 |
| 次のメッセージを追加しました。 02006-1 |

変更内容 (3020-6-375-10) HiRDB Text Search Plug-in Version 8 08-03

| 追加・変更内容 |
|-----------------------------------|
| 環境定義ファイルの定義例を変更しました。 |
| インデクス用グローバルバッファを最適化できるようにしました。 |
| FREEWORD 型でスコアオプション指定をできるようにしました。 |
| エラーコードで詳細化指定をできるようにしました。 |
| システムの制限を追加しました。 |
| 運用環境のバックアップを追加しました。 |
| SQL 連絡領域に出力されるメッセージを追加しました。 |

はじめに

このマニュアルは、次に示すプログラムプロダクトの機能、システム構築、環境設定および運用方法について説明したものです。

- P-1MD3-5381 HiRDB Text Search Plug-in Version 8(64)
- P-F1MD3-33711 HiRDB Text Search Plug-in Conceptual Extension Version 7
- P-9WD3-5381 HiRDB Text Search Plug-in Version 8(64)
- P-F9SD3-33711 HiRDB Text Search Plug-in Conceptual Extension Version 7
- P-24D3-3384 HiRDB Text Search Plug-in Version 8
- P-24D3-9374 HiRDB Text Search Plug-in Conceptual Extension Version 7

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 とします。

対象読者

このマニュアルは次の方にお読みいただくことを前提に説明しています。

- Text Search Plug-in のシステムを構築し、構築したシステムの環境を管理・運用する HiRDB 管理者の方
- HiRDB に関する知識および基本操作を習得している方
- UNIX または Windows に関する基本的な知識をお持ちの方
- SGML に関する基本的な知識をお持ちの方
- XML に関する基本的な知識をお持ちの方

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 解説

Text Search Plug-in の概要、特長、および機能について説明しています。

第 2 章 Text Search Plug-in システムの構築

Text Search Plug-in の運用を開始するために必要な定義ファイルの概要、Text Search Plug-in の環境設定およびデータベースの作成について説明しています。また、ユーザごとの環境設定方法についても説明しています。

第 3 章 Text Search Plug-in のユティリティ

Text Search Plug-in の運用に必要なユティリティの使用方法について説明しています。

第 4 章 抽象データ型関数を使用した文書または文字列データの操作

Text Search Plug-in で文書や文字列データを操作するために、Text Search Plug-in が提供している、抽象データ型関数の使用方法および検索条件の文法について説明しています。

第 5 章 推奨する運用方法

Text Search Plug-in での推奨する運用方法および設定情報について説明しています。

はじめに

- 付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字
登録する文書（テキストデータ）および文字列データに使用できる文字について説明しています。
- 付録 B ディレクトリおよびファイル構成
Text Search Plug-in を HiRDB に登録したあとの、ディレクトリおよびファイル構成を示しています。
- 付録 C 資源の排他
資源の排他制御について説明しています。
- 付録 D 障害発生時の運用方法
障害発生時の対処方法について説明しています。
- 付録 E 系切り替え
Text Search Plug-in の系切り替えについて説明しています。
- 付録 F エラーメッセージ
Text Search Plug-in で出力されるエラーメッセージの内容について説明しています。
- 付録 G 用語解説
このマニュアルで使用する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。

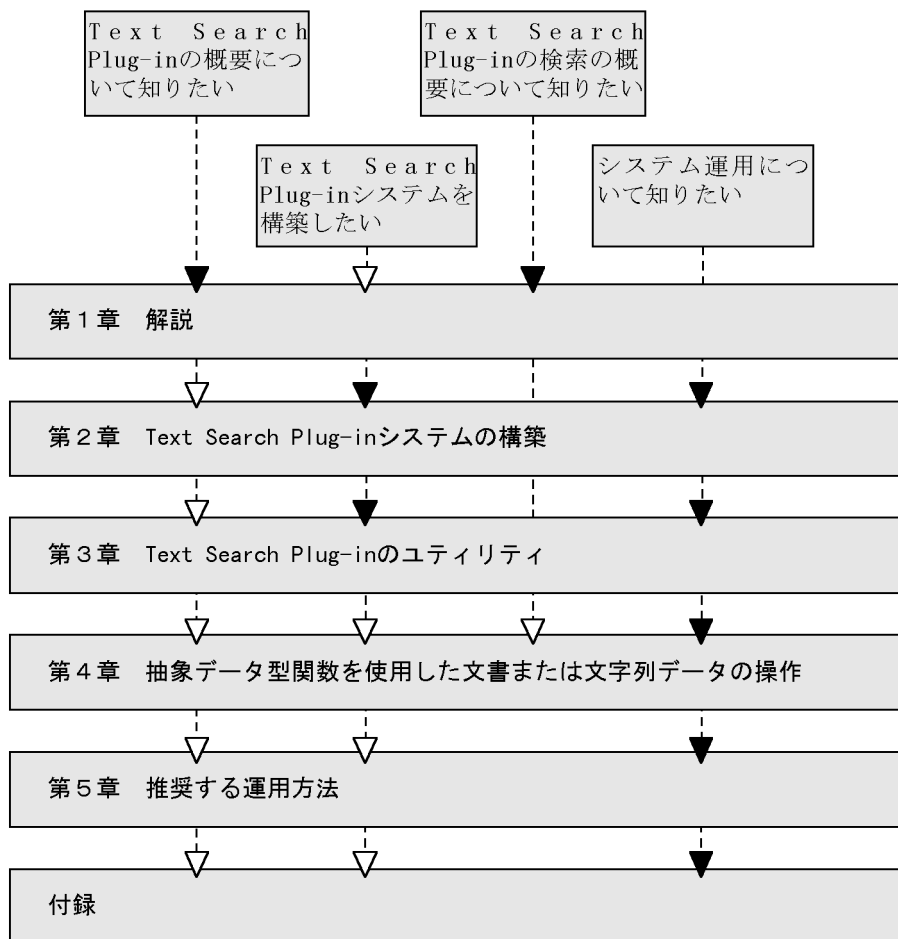
- HiRDB Version 8 解説（UNIX(R)用）(3000-6-351)
- HiRDB Version 8 システム導入・設計ガイド（UNIX(R)用）(3000-6-352)
- HiRDB Version 8 システム定義（UNIX(R)用）(3000-6-353)
- HiRDB Version 8 システム運用ガイド（UNIX(R)用）(3000-6-354)
- HiRDB Version 8 コマンドリファレンス（UNIX(R)用）(3000-6-355)
- HiRDB Version 8 解説（Windows(R)用）(3020-6-351)
- HiRDB Version 8 システム導入・設計ガイド（Windows(R)用）(3020-6-352)
- HiRDB Version 8 システム定義（Windows(R)用）(3020-6-353)
- HiRDB Version 8 システム運用ガイド（Windows(R)用）(3020-6-354)
- HiRDB Version 8 コマンドリファレンス（Windows(R)用）(3020-6-355)
- HiRDB Version 8 UAP 開発ガイド（UNIX(R)/Windows(R)用）(3020-6-356)
- HiRDB Version 8 SQL リファレンス（UNIX(R)/Windows(R)用）(3020-6-357)
- HiRDB Version 8 メッセージ（UNIX(R)/Windows(R)用）(3020-6-358)
- HiRDB Staticizer Option Version 8 (3000-6-363)
- HiRDB Version 8 XML Extension (3020-6-376)
- Preprocessing Library for Text Search Version 2 (3000-7-270)
- Text Search Enhancer for English (3000-6-301)

なお、このマニュアルでは、HiRDB 関連のマニュアルについて「スケーラブルデータベース

サーバ」を省略して表記しています。また、「(UNIX(R)用)」、「(UNIX(R)/Windows(R)用)」および「(Windows(R)用)」も省略して表記しています。

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。

●パーソナル
コンピュータまたは
ワークステーション



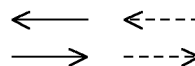
●入出力の動作



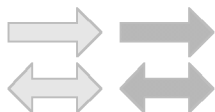
●ユーザ



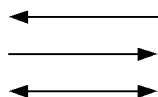
●制御の流れ



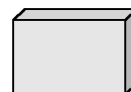
●データの流れ



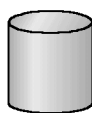
●その他の流れ



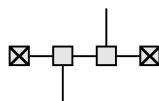
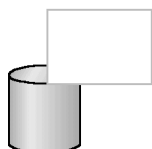
●文書またはファイル ●プログラム



●データベース



●データベースの内容 ●バス形のLAN



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 |
|------------|---|
| ストローク | 横に並べられた複数の項目に対し、項目間の区切りを示し、「または」の意味を示します。 (例) A B C は、「A、B または C」を意味します。 |
| { } 波括弧 | 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。 (例) { A B C } では、A、B または C のどれか一つを指定することを示します。 |
| [] きっ甲 | この記号で囲まれている項目は省略してもよいことを示します。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) [A] は「何も指定しない」か「A を指定する」ことを示します。 [B C] では「何も指定しない」か「B または C を指定する」ことを示します。 |
| < > 山括弧 | この記号で囲まれた複数の項目を一つの単位として、繰り返して指定できます。 (例) < A B C > では、次のように指定できることを示します。 A B C A B C |

| 記号 | 意味 |
|--------------|---|
| ... リーダー | 記述が省略されていることを示します。この記号の直前に示された項目を繰り返し複数個指定できます。 (例) A, B, B...では、「Aの後ろにBを複数個指定する」ことを示します。 |
| 白三角 | 1文字の半角スペースを示します。 |
| n 白三角とn文字 | n文字以上の半角スペースを示します。 |

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 | ! " # \$ % & ' () + , - . / : ; < = > @ [] ^ _ { } タブ スペース |

注

ただし、かたかな文字の直後にある「-」は、かたかな文字になります。

このマニュアルで使用する計算式の記号

このマニュアルで使用する計算式の記号の意味を次に示します。

| 記号 | 意味 |
|-----|---|
| | 計算結果の値の小数点以下を切り上げることを示します。 (例) $34 \div 3$ の計算結果は 12 となります。 |
| | 計算結果の値の小数点以下を切り捨てることを示します。 (例) $34 \div 3$ の計算結果は 11 となります。 |
| max | 計算結果の最も大きい値を選ぶことを示します。 (例) $\max(3 \times 6, 4 + 7)$ の計算結果は 18 となります。 |
| min | 計算結果の最も小さい値を選ぶことを示します。 (例) $\min(3 \times 6, 4 + 7)$ の計算結果は 11 となります。 |

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 | <u>A</u> pplication <u>P</u> rogram |
| 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 |
| DB | <u>D</u> ata <u>B</u> ase |
| DIC | <u>D</u> ictionary <u>S</u> erver |
| DTD | <u>D</u> ocument <u>T</u> ype <u>D</u> efinition |
| EUC | <u>E</u> xtended <u>U</u> NIX <u>C</u> ode |
| FES | <u>F</u> ront <u>E</u> nd <u>S</u> erver |
| FTP | <u>F</u> ile <u>T</u> ransfer <u>P</u> rotocol |
| IPF | <u>I</u> tanium(R) <u>P</u> rocessor <u>F</u> amily |
| JIS | <u>J</u> apanese <u>I</u> ndustrial <u>S</u> tandard code |
| MGR | <u>S</u> ystem <u>M</u> anager |
| OS | <u>O</u> perating <u>S</u> ystem |
| RDB | <u>R</u> elational <u>D</u> ata <u>B</u> ase |
| SDS | <u>S</u> ingle <u>D</u> atabase <u>S</u> erver |
| SGML | <u>S</u> tandard <u>G</u> eneralized <u>M</u> arkup <u>L</u> anguage |
| SQL | <u>S</u> tructured <u>Q</u> uery <u>L</u> anguage |
| UAP | <u>U</u> ser <u>A</u> pplication <u>P</u> rogram |
| XML | <u>e</u> Xtensible <u>M</u> arkup <u>L</u> anguage |

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 | |
|--|--------------------------------|-------|
| AIX 5L V5.1 | AIX | |
| AIX 5L V5.2 | | |
| AIX 5L V5.3 | | |
| AIX V6.1 | | |
| HiRDB/Single Server Version 8 HiRDB/Single Server Version 8 | HiRDB/ シングルサーバ (またはシングルサーバ) | HiRDB |
| HiRDB/Parallel Server Version 8 HiRDB/Parallel Server Version 8 | HiRDB/ パラレルサーバ (またはパラレルサーバ) | |

| 製品名称 | 略称 | |
|---|--|---------|
| HiRDB/Developer's Kit Version 8 および HiRDB/Run Time Version 8 HiRDB/Developer's Kit Version 8 および HiRDB/Run Time Version 8 | HiRDB クライアント機能 | |
| HiRDB/Single Server - Object Option および HiRDB/Parallel Server - Object Option | Object Option | |
| HiRDB Staticizer Option | Staticizer Option | |
| HiRDB Text Search Plug-in Version 8 | Text Search Plug-in | |
| HiRDB Text Search Plug-in Conceptual Extension Version 7 | Text Search Plug-in Conceptual Extension | |
| 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(R) XP Professional Operating System | Windows XP または Windows XP Professional | |
| Microsoft(R) Windows(R) XP Professional x64 Edition | Windows XP または Windows XP x64 Edition | |
| Microsoft(R) Windows Server(R) 2003, Standard Edition | Windows Server 2003 | |
| Microsoft(R) Windows Server(R) 2003, Enterprise Edition | | |
| Microsoft(R) Windows Server(R) 2003, Standard x64 Edition | Windows Server 2003 または | |
| Microsoft(R) Windows Server(R) 2003, Enterprise x64 Edition | Windows Server 2003 x64 Editions | |
| Microsoft(R) Windows Vista(R) Business Edition | Windows Vista | |
| Microsoft(R) Windows Vista(R) Enterprise Edition | | |
| Microsoft(R) Windows Vista(R) Ultimate Edition | | |
| Microsoft(R) Windows Vista(R) Business(x64) Edition | | |
| Microsoft(R) Windows Vista(R) Enterprise(x64) Edition | | |
| Microsoft(R) Windows Server(R) 2008, Standard | Windows Server 2008 | |
| Microsoft(R) Windows Server(R) 2008, Enterprise | | |
| Microsoft(R) Windows Server(R) 2008, Standard(x64) | | |
| Microsoft(R) Windows Server(R) 2008, Enterprise(x64) | | |
| Preprocessing Library for Text Search および Preprocessing Library for Text Search Version 2 | Preprocessing Library for Text Search | |

はじめに

- 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「/」で表記している箇所があります。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。

個所（かしょ） 必須（ひつす） 閉塞（へいそく） 拗音（ようおん）

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、 $1,024^2$ バイト、 $1,024^3$ バイト、 $1,024^4$ バイトです。

目次

| | | |
|----------|--|-----------|
| 1 | 解説 | 1 |
| 1.1 | Text Search Plug-in とは | 2 |
| 1.1.1 | Text Search Plug-in の機能 | 2 |
| 1.1.2 | Text Search Plug-in の特長 | 2 |
| 1.2 | Text Search Plug-in のシステム構成 | 4 |
| 1.2.1 | Text Search Plug-in を構成するプログラム | 4 |
| 1.2.2 | Text Search Plug-in システムの構成例 | 5 |
| 1.3 | Text Search Plug-in を利用した検索 | 7 |
| 1.3.1 | 全文検索 | 7 |
| 1.3.2 | 概念検索 | 9 |
| 1.4 | Text Search Plug-in で効果的に検索するには | 14 |
| 2 | Text Search Plug-in システムの構築 | 17 |
| 2.1 | Text Search Plug-in システムの構築の流れ | 18 |
| 2.2 | システムの設計 | 20 |
| 2.2.1 | 検討項目 | 20 |
| 2.2.2 | 使用する文字コードによる機能差異について | 22 |
| 2.3 | リソースの見積もり | 26 |
| 2.3.1 | メモリ所要量 | 26 |
| 2.3.2 | RD エリア容量 | 31 |
| 2.3.3 | ファイル容量 | 41 |
| 2.3.4 | 論理ファイル数 | 47 |
| 2.4 | Text Search Plug-in の環境設定 | 49 |
| 2.4.1 | Text Search Plug-in のセットアップ (UNIX の場合) | 49 |
| 2.4.2 | Text Search Plug-in の登録 | 50 |
| 2.4.3 | レジストリ機能の初期設定 | 52 |
| 2.4.4 | システム共通定義の変更 | 53 |
| 2.5 | 文書検索プラグインのデータベースの作成 | 55 |
| 2.5.1 | 文書検索プラグインのデータベース作成の流れ | 55 |
| 2.5.2 | SGML 定義情報の作成 | 56 |
| 2.5.3 | SGML 定義情報の登録 | 62 |
| 2.5.4 | データ格納領域の確保 | 63 |
| 2.5.5 | 表の作成 | 65 |

| | | |
|--------|--|-----|
| 2.5.6 | インデクスの定義 | 67 |
| 2.5.7 | データの登録 | 85 |
| 2.5.8 | 同義語辞書の作成 | 88 |
| 2.5.9 | 同義語辞書の登録 | 90 |
| 2.5.10 | SGML 定義情報のメンテナンス | 90 |
| 2.6 | 文書検索プラグインのユーザ環境設定 | 92 |
| 2.6.1 | SGML プラグインの環境設定 | 92 |
| 2.6.2 | n-gram インデクスプラグインの環境設定 | 95 |
| 2.7 | 文字列検索プラグインのデータベースの作成 | 104 |
| 2.7.1 | 文字列検索プラグインのデータベース作成の流れ | 104 |
| 2.7.2 | データ格納領域の確保 | 105 |
| 2.7.3 | 表の作成 | 105 |
| 2.7.4 | インデクスの定義 | 106 |
| 2.7.5 | データの登録 | 120 |
| 2.7.6 | 同義語辞書の作成 | 123 |
| 2.7.7 | 同義語辞書の登録 | 123 |
| 2.8 | 文字列検索プラグインのユーザ環境設定 | 124 |
| 2.8.1 | FREEWORD プラグインの環境設定 | 124 |
| 2.8.2 | IXFREEWORD プラグインの環境設定 | 126 |
| 2.9 | システムの制限 | 134 |
| 2.10 | Text Search Plug-in の設定解除 | 135 |
| 2.10.1 | 設定解除の流れ | 135 |
| 2.10.2 | 表の削除 | 136 |
| 2.10.3 | レジストリ情報の削除 (SGMLTEXT 型の列を定義していた場合) | 136 |
| 2.10.4 | Text Search Plug-in の削除 | 136 |
| 2.10.5 | Text Search Plug-in のアンセットアップ (UNIX の場合) | 138 |
| 2.10.6 | システム共通定義の変更 | 139 |
| 2.10.7 | HiRDB を OS から削除する場合の注意事項 (UNIX の場合) | 139 |
| 2.11 | HiRDB の提供する機能の利用 | 141 |
| <hr/> | | |
| 3 | Text Search Plug-in のユティリティ | 143 |
| 3.1 | ユティリティの使用方法 | 144 |
| 3.1.1 | Text Search Plug-in のユティリティとは | 144 |
| 3.1.2 | ユティリティの記述形式と操作 | 147 |
| 3.2 | 文書検索プラグインのユティリティ | 150 |
| 3.2.1 | phsregsetup (レジストリ初期化ユティリティ) | 150 |

| | | |
|----------|---|------------|
| 3.2.2 | phssgmlreg (SGML 定義情報登録ユティリティ) | 151 |
| 3.2.3 | phssgmlrm (SGML 定義情報削除ユティリティ) | 152 |
| 3.2.4 | phssgmlval (SGML 定義情報参照ユティリティ) | 153 |
| 3.2.5 | phssyndicreg (同義語辞書登録ユティリティ) | 154 |
| 3.2.6 | phssyndicrm (同義語辞書削除ユティリティ) | 156 |
| 3.2.7 | phssyndicconv (同義語辞書移行ユティリティ) | 157 |
| 3.2.8 | phssyndicbkup (同義語辞書バックアップユティリティ)(UNIX の場合) | 158 |
| 3.2.9 | phnidxls (インデクス情報の取得ユティリティ) | 159 |
| 3.2.10 | phnincmt (インデクスのインクリメンタルユティリティ) | 163 |
| 3.2.11 | phnmerge (インデクスのマージユティリティ) | 166 |
| 3.2.12 | phnexpnd (インデクスの拡張ユティリティ) | 168 |
| 3.2.13 | phnstartidx (インデクス作成開始ユティリティ) | 170 |
| 3.2.14 | phndelay (インデクス分割遅延登録運用変更ユティリティ) | 173 |
| 3.2.15 | phnmodidx (インデクス情報変更ユティリティ) | 175 |
| 3.2.16 | phncond (インデクス詰め替えユティリティ) | 178 |
| 3.2.17 | phnconget (統計情報収集ユティリティ) | 181 |
| 3.2.18 | phnconbuild (統計情報統合ユティリティ) | 185 |
| 3.2.19 | phnconstr (統計情報反映ユティリティ) | 187 |
| 3.2.20 | phnchgcon (特徴ターム取得 RD エリア指定ユティリティ) | 191 |
| 3.3 | 文字列検索プラグインのユティリティ | 194 |
| 3.3.1 | phcsyndicreg (同義語辞書登録ユティリティ) | 194 |
| 3.3.2 | phcsyndicrm (同義語辞書削除ユティリティ) | 196 |
| 3.3.3 | phcsyndicconv (同義語辞書移行ユティリティ) | 197 |
| 3.3.4 | phcsyndicbkup (同義語辞書バックアップユティリティ)(UNIX の場合) | 198 |
| 3.3.5 | phxidxls (インデクス情報の取得ユティリティ) | 199 |
| 3.3.6 | phxincmt (インデクスのインクリメンタルユティリティ) | 202 |
| 3.3.7 | phxexpnd (インデクスの拡張ユティリティ) | 205 |
| 3.3.8 | phxstartidx (インデクス作成開始ユティリティ) | 207 |
| 3.3.9 | phxdelay (インデクス分割遅延登録運用変更ユティリティ) | 211 |
| 3.3.10 | phxmodidx (インデクス情報変更ユティリティ) | 213 |
| 3.3.11 | phxcond (インデクス詰め替えユティリティ) | 216 |
| 4 | 抽象データ型関数を使用した文書または文字列データの操作 | 219 |
| 4.1 | Text Search Plug-in が提供する抽象データ型関数 | 220 |
| 4.1.1 | 抽象データ型関数の使用方法 | 220 |
| 4.1.2 | 抽象データ型の選択 | 222 |

| | | |
|--------|---------------------------|-----|
| 4.2 | 抽象データ型関数の文法 | 224 |
| 4.2.1 | SGMLTEXT | 224 |
| 4.2.2 | FREEWORD | 226 |
| 4.2.3 | contains | 227 |
| 4.2.4 | contains_with_score | 228 |
| 4.2.5 | score | 230 |
| 4.2.6 | extracts | 232 |
| 4.2.7 | concept_with_score | 237 |
| 4.2.8 | score_concept | 239 |
| 4.2.9 | concept_with_terms | 241 |
| 4.2.10 | terms_concept | 243 |
| 4.2.11 | 検索条件式の指定に関する注意 | 246 |
| 4.3 | 全文検索条件式の文法（文書検索プラグインの場合） | 247 |
| 4.3.1 | 構造名を指定した検索条件の AND/OR 論理条件 | 247 |
| 4.3.2 | 構造名の指定 | 248 |
| 4.3.3 | 構造名のワイルドカード指定 | 249 |
| 4.3.4 | テキスト検索条件の指定 | 250 |
| 4.3.5 | 単純文字列指定 | 251 |
| 4.3.6 | 異表記展開指定 | 254 |
| 4.3.7 | 同義語展開指定 | 257 |
| 4.3.8 | 近傍条件指定 | 258 |
| 4.3.9 | 重み | 260 |
| 4.3.10 | 属性値指定 | 261 |
| 4.3.11 | 特定構造検索指定 | 261 |
| 4.3.12 | 特殊文字の指定 | 262 |
| 4.3.13 | 全文検索条件式の指定例 | 263 |
| 4.4 | 全文検索条件式の文法（文字列検索プラグインの場合） | 266 |
| 4.4.1 | 検索条件の AND/OR 論理条件 | 266 |
| 4.4.2 | テキスト検索条件の指定 | 267 |
| 4.4.3 | 単純文字列指定 | 268 |
| 4.4.4 | 異表記展開指定 | 269 |
| 4.4.5 | 同義語展開指定 | 272 |
| 4.4.6 | 近傍条件指定 | 272 |
| 4.4.7 | 特殊文字の指定 | 274 |
| 4.4.8 | 全文検索条件式の指定例 | 274 |
| 4.5 | 概念検索条件式の文法 | 276 |
| 4.5.1 | 構造名を指定した検索条件の AND/OR 論理条件 | 276 |

| | | |
|--------|---------------|-----|
| 4.5.2 | 構造名の指定 | 277 |
| 4.5.3 | 構造名のワイルドカード指定 | 277 |
| 4.5.4 | テキスト検索条件の指定 | 279 |
| 4.5.5 | 単純文字列指定 | 280 |
| 4.5.6 | 異表記展開指定 | 281 |
| 4.5.7 | 同義語展開指定 | 282 |
| 4.5.8 | 重み | 283 |
| 4.5.9 | スコアオプション | 283 |
| 4.5.10 | 検索オプション | 285 |
| 4.5.11 | 絞り込み指定 | 285 |
| 4.5.12 | 概念検索条件式の指定例 | 286 |

5

| | |
|--------------------------|-----|
| 推奨する運用方法 | 289 |
| 5.1 推奨するデータ登録の運用方法 | 290 |
| 5.1.1 登録運用の種類 | 290 |
| 5.1.2 推奨する登録運用 | 293 |
| 5.2 性能向上の運用 | 298 |
| 5.2.1 グローバルバッファ管理の運用 | 298 |
| 5.3 ログおよびバックアップの取得方式 | 299 |
| 5.3.1 登録運用のログ取得方式の設定について | 299 |
| 5.3.2 バックアップ取得方法 | 299 |
| 5.4 概念検索用統計情報の収集，統合および反映 | 302 |

付録

| | |
|---------------------------------------|-----|
| 付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字 | 304 |
| 付録 B ディレクトリおよびファイル構成 | 307 |
| 付録 B.1 ディレクトリおよびファイル構成（UNIX の場合） | 307 |
| 付録 B.2 ディレクトリおよびファイル構成（Windows の場合） | 312 |
| 付録 C 資源の排他 | 318 |
| 付録 C.1 排他取得内容 | 318 |
| 付録 C.2 レジストリの排他 | 321 |
| 付録 C.3 デッドロックの防止 | 321 |
| 付録 D 障害発生時の運用方法 | 322 |
| 付録 E 系切り替え | 323 |
| 付録 F エラーメッセージ | 325 |

| | |
|-----------------------------------|-----|
| 付録 F.1 SQL 連絡領域に出力されるメッセージ | 325 |
| 付録 F.2 メッセージログ・標準エラー出力に出力されるメッセージ | 327 |
| 付録 F.3 メッセージの記述形式 | 327 |
| 付録 F.4 メッセージ一覧 | 328 |
| 付録 F.5 SGML 文書の登録エラーメッセージ | 373 |
| 付録 F.6 SGML 文書の登録エラーメッセージ一覧 | 375 |
| 付録 G 用語解説 | 385 |

| | |
|-----------|------------|
| 索引 | 391 |
|-----------|------------|

目次

| | | |
|--------|---|----|
| 図 1-1 | Text Search Plug-in システムの構成例 | 6 |
| 図 1-2 | Text Search Plug-in での全文検索の概要 | 7 |
| 図 1-3 | Text Search Plug-in Conceptual Extension を組み込んだ場合の概念検索の概要 | 11 |
| 図 2-1 | Text Search Plug-in システムの構築の流れ | 18 |
| 図 2-2 | データベース作成の流れ（文書検索プラグイン） | 56 |
| 図 2-3 | DTD と SGML 文書の分割 | 58 |
| 図 2-4 | 正規化パラメタファイルの記述例 | 61 |
| 図 2-5 | SGML 文書の正規化の例 | 62 |
| 図 2-6 | create rdarea 文の例（UNIX の場合） | 64 |
| 図 2-7 | create rdarea 文の例（Windows の場合） | 65 |
| 図 2-8 | 表の定義例（SGML 文書または XML 文書の場合） | 67 |
| 図 2-9 | 表の定義例（プレーン文書の場合） | 67 |
| 図 2-10 | 表の定義例（XML 文書で utf-8（MS-Unicode）の場合） | 67 |
| 図 2-11 | 全文検索機能使用時のインデクス定義例 | 69 |
| 図 2-12 | 概念検索機能使用時のインデクス定義例 | 69 |
| 図 2-13 | インデクスファイル自動拡張機能使用時のインデクス定義例（全文検索用インデクスの場合） | 70 |
| 図 2-14 | インデクス作成ルール追加機能使用時（INDEX_MEDIUM）のインデクス定義例（全文検索用のインデクスの場合） | 71 |
| 図 2-15 | インデクス作成ルール追加機能使用時（INDEX_ALL）のインデクス定義例（概念検索用のインデクスの場合） | 71 |
| 図 2-16 | 大小文字統一機能使用時のインデクス定義例（全文検索用のインデクスの場合） | 72 |
| 図 2-17 | 全角文字半角文字統一機能使用時のインデクス定義例（全文検索用のインデクスの場合） | 72 |
| 図 2-18 | 大小文字統一機能，および全角文字半角文字統一機能使用時のインデクス定義例（概念検索用のインデクスの場合） | 72 |
| 図 2-19 | 拗音統一機能使用時のインデクス定義例 | 73 |
| 図 2-20 | 濁音統一機能使用時のインデクス定義例 | 74 |
| 図 2-21 | 削除文字指定機能使用時のインデクス定義例 | 74 |
| 図 2-22 | 文字単位インデクス作成機能使用時のインデクス定義例 | 75 |
| 図 2-23 | 差分インデクス使用時のインデクス定義例（全文検索用のインデクスの場合） | 76 |
| 図 2-24 | 差分インデクス使用時のインデクス定義例（概念検索用のインデクスの場合） | 76 |
| 図 2-25 | 初期分割登録使用時のインデクス定義例（全文検索用のインデクスの場合） | 78 |
| 図 2-26 | 初期分割登録使用時のインデクス定義例（概念検索用のインデクスの場合） | 78 |

| | | |
|--------|---------------------------------------|-----|
| 図 2-27 | 分割遅延登録使用時のインデクス定義例（全文検索用のインデクスの場合） | 79 |
| 図 2-28 | 分割遅延登録使用時のインデクス定義例（概念検索用のインデクスの場合） | 79 |
| 図 2-29 | 分割遅延登録への移行時に実行する，インデクス情報変更ユーティリティの指定例 | 79 |
| 図 2-30 | 分割遅延登録・更新系 SQL 同時実行使用時のインデクス定義例 | 80 |
| 図 2-31 | 英文検索機能を使用する場合のインデクス定義例（英文検索） | 82 |
| 図 2-32 | INSERT 文の処理の例 | 85 |
| 図 2-33 | 同義語辞書テキストファイルの記述例 | 89 |
| 図 2-34 | 同義語のグループでの文字種の分割例 | 90 |
| 図 2-35 | SGML プラグインの環境定義ファイルの定義例 | 93 |
| 図 2-36 | n-gram インデクスプラグインの環境定義ファイルの定義例 | 97 |
| 図 2-37 | データベース作成の流れ（文字列検索プラグイン） | 104 |
| 図 2-38 | create rdarea 文の例（UNIX の場合） | 105 |
| 図 2-39 | create rdarea 文の例（Windows の場合） | 105 |
| 図 2-40 | 表の定義例 | 106 |
| 図 2-41 | 全文検索機能使用時のインデクス定義例 | 107 |
| 図 2-42 | インデクスファイル自動拡張機能使用時のインデクス定義例 | 107 |
| 図 2-43 | 大小文字統一機能使用時のインデクス定義例 | 108 |
| 図 2-44 | 全角文字半角文字統一機能使用時のインデクス定義例 | 108 |
| 図 2-45 | 大小文字統一機能，および全角文字半角文字統一機能使用時のインデクス定義例 | 108 |
| 図 2-46 | 拗音統一機能使用時のインデクス定義例 | 109 |
| 図 2-47 | 濁音統一機能使用時のインデクス定義例 | 110 |
| 図 2-48 | 削除文字指定機能使用時のインデクス定義例 | 110 |
| 図 2-49 | 文字単位インデクス作成機能使用時のインデクス定義例 | 111 |
| 図 2-50 | 初期分割登録使用時のインデクス定義例 | 112 |
| 図 2-51 | 分割遅延登録使用時のインデクス定義例 | 113 |
| 図 2-52 | 分割遅延登録への移行時に実行する，インデクス情報変更ユーティリティの指定例 | 113 |
| 図 2-53 | 分割遅延登録・更新系 SQL 同時実行のインデクスの定義例 | 114 |
| 図 2-54 | 英文検索機能を使用する場合のインデクス定義例（英語文字列検索） | 116 |
| 図 2-55 | 登録件数指定のインデクス定義例 | 120 |
| 図 2-56 | INSERT 文の処理の例 | 121 |
| 図 2-57 | FREEWORD プラグインの環境定義ファイルの定義例 | 125 |
| 図 2-58 | IXFREEWORD プラグインの環境定義ファイルの定義例 | 128 |
| 図 2-59 | Text Search Plug-in の設定解除の流れ | 135 |
| 図 3-1 | 統計情報出力先定義ファイルの定義例（UNIX の場合） | 183 |
| 図 3-2 | 統計情報出力先定義ファイルの定義例（Windows の場合） | 183 |

| | | |
|--------|---------------------------------------|-----|
| 図 3-3 | 入力統計情報ファイル格納先定義ファイルの定義例 (UNIX の場合) | 186 |
| 図 3-4 | 入力統計情報ファイル格納先定義ファイルの定義例 (Windows の場合) | 186 |
| 図 3-5 | 統計情報定義ファイルの定義例 (UNIX の場合) | 189 |
| 図 3-6 | 統計情報定義ファイルの定義例 (Windows の場合) | 189 |
| 図 4-1 | 検索用特徴タームと重みの取得形式 | 243 |
| 図 4-2 | 全文検索条件式の形式 | 247 |
| 図 4-3 | 構造検索パターンの形式 | 248 |
| 図 4-4 | 構造名指定の形式 | 249 |
| 図 4-5 | テキスト検索条件の形式 | 250 |
| 図 4-6 | 検索因子の形式 | 251 |
| 図 4-7 | 単純文字列指定の形式 | 251 |
| 図 4-8 | 異表記展開指定の形式 | 254 |
| 図 4-9 | 同義語展開指定の形式 | 258 |
| 図 4-10 | 近傍条件指定の形式 | 259 |
| 図 4-11 | 英文インデクスの近傍条件指定の形式 | 260 |
| 図 4-12 | 重み指定の形式 | 260 |
| 図 4-13 | 属性値指定の形式 | 261 |
| 図 4-14 | 特定構造検索指定の形式 | 261 |
| 図 4-15 | 特定構造指定の形式 | 262 |
| 図 4-16 | 全文検索条件式の形式 | 266 |
| 図 4-17 | テキスト検索条件の形式 | 267 |
| 図 4-18 | 検索因子の形式 | 267 |
| 図 4-19 | 単純文字列指定の形式 | 268 |
| 図 4-20 | 異表記展開指定の形式 | 269 |
| 図 4-21 | 同義語展開指定の形式 | 272 |
| 図 4-22 | 近傍条件指定の形式 | 273 |
| 図 4-23 | 英文インデクスの近傍条件指定の形式 | 274 |
| 図 4-24 | 概念検索条件式の形式 | 276 |
| 図 4-25 | 構造検索パターンの形式 | 277 |
| 図 4-26 | 構造名指定の形式 | 278 |
| 図 4-27 | テキスト検索条件の形式 | 279 |
| 図 4-28 | 検索因子の形式 | 279 |
| 図 4-29 | 単純文字列指定の形式 | 280 |
| 図 4-30 | 単純文字列指定の形式 (検索オプションに term を指定した場合) | 280 |
| 図 4-31 | 異表記展開指定の形式 | 281 |

| | | |
|--------|---------------|-----|
| 図 4-32 | 同義語展開指定の形式 | 282 |
| 図 4-33 | 重み指定の形式 | 283 |
| 図 4-34 | スコアオプション指定の形式 | 284 |
| 図 4-35 | 検索オプション指定の形式 | 285 |

表目次

| | | |
|--------|--|-----|
| 表 2-1 | 使用できる検索機能 | 22 |
| 表 2-2 | インデクス作成時に使用できる機能 | 23 |
| 表 2-3 | 使用できるユティリティ | 23 |
| 表 2-4 | 文書検索プラグインの SGML 定義情報の作成に必要なファイル | 57 |
| 表 2-5 | SGML 文書中の改行コードの変換方法 | 59 |
| 表 2-6 | SGML 定義情報の指定（文字コードが Shift-JIS または EUC の場合） | 66 |
| 表 2-7 | SGML 定義情報の指定（文字コードが utf-8 の場合） | 66 |
| 表 2-8 | CREATE INDEX のプラグインオプションに対する定義内容 | 70 |
| 表 2-9 | 定義内容 | 71 |
| 表 2-10 | SGML プラグインの環境定義ファイルに設定できる定義項目 | 93 |
| 表 2-11 | n-gram インデクスプラグインの環境定義ファイルに設定できる定義項目 | 97 |
| 表 2-12 | 定義内容 | 107 |
| 表 2-13 | FREWORD プラグインの環境定義ファイルに設定できる定義項目 | 125 |
| 表 2-14 | IXFREWORD プラグインの環境定義ファイルに設定できる定義項目 | 128 |
| 表 2-15 | コンストラクタ関数の情報 | 141 |
| 表 2-16 | コンストラクタパラメタ逆生成関数 | 141 |
| 表 3-1 | Text Search Plug-in のユティリティの機能一覧 | 144 |
| 表 3-2 | 同義語辞書名に指定できない記号および英数字 | 155 |
| 表 4-1 | Text Search Plug-in の抽象データ型関数一覧 | 221 |
| 表 4-2 | 抽象データ型の相違点 | 222 |
| 表 4-3 | スコアの算出方法 | 231 |
| 表 4-4 | 指定できるワイルドカードおよび特殊文字 | 252 |
| 表 4-5 | 指定できるワイルドカードおよび特殊文字（英文の場合） | 252 |
| 表 4-6 | 指定できる除外文字検索用の特殊文字 | 253 |
| 表 4-7 | 異表記オプション | 254 |
| 表 4-8 | 全文検索条件の条件式中に指定できる特殊記号（全文検索条件の単純文字列指定以外） | 262 |
| 表 4-9 | 全文検索条件の条件式中に指定できる特殊記号（全文検索条件の単純文字列指定内） | 263 |
| 表 4-10 | 異表記オプション | 269 |
| 表 4-11 | 異表記オプション | 281 |
| 表 5-1 | Text Search Plug-in の登録運用の長所および短所 | 291 |
| 表 5-2 | 登録・検索同時実行による運用の長所および短所 | 292 |

| | | |
|-------|---|-----|
| 表 5-3 | 登録運用で推奨するパラメタの組み合わせ | 297 |
| 表 A-1 | 登録する文書（テキストデータ）および文字列データに使用できる文字（HiRDB の文字コード種別が SHIFT-JIS のとき） | 304 |
| 表 A-2 | 登録する文書（テキストデータ）および文字列データに使用できる文字（HiRDB の文字コード種別が EUC のとき）(UNIX の場合) | 304 |
| 表 A-3 | 登録する文書（テキストデータ）および文字列データに使用できる文字（HiRDB の文字コード種別が SHIFT-JIS , EUC および utf-8 のとき） | 305 |
| 表 A-4 | utf-8 コードで使用できる文字（表内の文字コード表現は 16 進数） | 305 |
| 表 B-1 | SGML プラグインと n-gram インデクスプラグインのディレクトリおよびファイル構成（UNIX の場合） | 307 |
| 表 B-2 | FREWORD プラグインと IXFREWORD プラグインのディレクトリおよびファイル構成（UNIX の場合） | 310 |
| 表 B-3 | SGML プラグインと n-gram インデクスプラグインのフォルダおよびファイル構成（Windows の場合） | 312 |
| 表 B-4 | FREWORD プラグインと IXFREWORD プラグインのフォルダおよびファイル構成（Windows の場合） | 315 |
| 表 C-1 | SQL 実行時の排他制御のモード（SQL に対応した論理ファイルの使用） | 318 |
| 表 C-2 | ユティリティコマンドの排他制御のモード（SGML プラグインおよび n-gram インデクスプラグイン） | 319 |
| 表 C-3 | SQL 実行時の排他制御のモード（SQL に対応した論理ファイルの使用） | 320 |
| 表 C-4 | ユティリティコマンドの排他制御のモード | 320 |
| 表 F-1 | SQL コード一覧 | 326 |
| 表 F-2 | SGML 文書の登録エラーメッセージ一覧 | 375 |

1

解説

この章では、Text Search Plug-in の概要、特長、および機能について説明します。

1.1 Text Search Plug-in とは

1.2 Text Search Plug-in のシステム構成

1.3 Text Search Plug-in を利用した検索

1.4 Text Search Plug-in で効果的に検索するには

1.1 Text Search Plug-in とは

企業やさまざまな研究機関では、知的財産として大量の文書を所有しています。しかし、単にデータベースに蓄積されているだけでは、知的財産としての役割を十分に果たせません。蓄積した文書を資産として利用するには、大量の文書から目的の情報をすばやく、しかも正確に抽出できるような仕組みが必要です。

Text Search Plug-in は、この「必要な情報をすばやく、正確に検索する」手段を提供します。

1.1.1 Text Search Plug-in の機能

Text Search Plug-in は、HiRDB に組み込んで使用するプログラムです。HiRDB のデータベースから、必要な情報を正確に、しかも高速に検索する手段として、次の二つの機能を提供しています。

文書（テキストデータ）を対象にした文書検索

文字列データを対象にした文字列検索

これらの機能を利用することで、任意のキーワード（検索ターム）を指定して文書や文字列データを検索できます（全文検索）。例えば、新聞記事を管理しているデータベースから、「情報技術」に関する記事を検索する場合、キーワードとして「情報技術」を指定して検索します。すると、「情報技術」という単語を含む記事がすべて検索できます。

また、新聞記事を SGML 文書で管理しているテキストデータベースから、「記事の見出しに『情報技術』という単語を含む記事を探す」というような、文書の構造を検索条件に指定した検索もできます。

さらに、Text Search Plug-in の関連プログラムである Text Search Plug-in Conceptual Extension を組み込むことで、キーワードを文章（種文章）で指定し、その文章と似た意味を持つ文書を検索できます（概念検索）。この検索方法では、検索条件のキーワードが限定できない場合や、複雑な検索条件の入力などの操作に不慣れな場合でも、文章の持つ概念で文書を検索できます。

1.1.2 Text Search Plug-in の特長

Text Search Plug-in には、次の特長があります。

（1）大量の文書または文字列データの管理

Text Search Plug-in で扱うデータには、文書（テキストデータ）の場合と、文字列データの場合があります。これらのデータは、HiRDB のデータベースに格納します。したがって、HiRDB の特長を生かし、高い検索性能や高い信頼性を持つシステムとして大量の文書または文字列データを管理できます。

(2) SGML や XML で書かれた構造を持つ文書の登録と検索

Text Search Plug-in では、扱うデータが文書の場合、SGML や XML のように文書の構造を定義する言語で書かれたファイルを文書として登録できます。検索条件には、文書に定義した構造を利用できます。例えば、「文書中の、『見出し』として定義された構造内の文章中に『情報技術』という単語が含まれる文書を探す」などの検索ができます。

ただし、XML 文書の場合は、Preprocessing Library for Text Search を利用して Text Search Plug-in で使用できるデータ形式に変換する必要があります。Preprocessing Library for Text Search でデータを登録する方法については、マニュアル「Preprocessing Library for Text Search Version 2」を参照してください。

(3) 多彩な検索方法

Text Search Plug-in を利用した検索には、単語をキーワードにして検索する全文検索と、任意の文章または文字列から似た内容の文章を検索する概念検索とがあります。さらに、これら全文検索または概念検索に、さまざまな機能を追加することで、多様な検索を実現できます。

例えば、検索タームと意味が同じで表記の異なる情報を検索したり（同義語・異表記展開検索）、検索条件を複数指定して、その条件の両方に一致する情報を検索したり（論理演算検索）できます。

1.2 Text Search Plug-in のシステム構成

この節では、Text Search Plug-in を構成するプログラムについて説明します。また、Text Search Plug-in システム構成例も示します。

1.2.1 Text Search Plug-in を構成するプログラム

Text Search Plug-in を構成するプログラムには、文書の操作を実現する文書検索プラグインと、文字列データの操作を実現する文字列検索プラグインがあります。

文書検索プラグインは、SGMLTEXT 型という抽象データ型で文書を扱います。また、文字列検索プラグインは、FREEWORD 型という抽象データ型で文字列データを扱います。文書と文字列データのどちらの抽象データ型を使用するかについては、「4.1.2 抽象データ型の選択」を参照してください。

各プログラムの概要を次に説明します。

(1) 文書検索プラグイン

文書検索プラグインは、SGML プラグインと n-gram インデクスプラグインで構成されています。

SGML プラグイン

SGML プラグインは、データ型を追加するプラグインです。このプラグインは、SQL 抽象データ型を前提としており、SGMLTEXT 型という抽象データ型を追加して、SGML 文書の検索、登録などの操作機能を実現します。

Text Search Plug-in での SGML プラグインの名称は、_phsgml です。

n-gram インデクスプラグイン

n-gram インデクスプラグインは、SGMLTEXT 型に対する検索インデクスを追加するプラグインで、NGRAM 型というインデクス型を追加します。n-gram インデクスプラグインは n-gram インデクス方式の検索エンジンを持ち、SGML プラグインに対して検索手段を提供します。

Text Search Plug-in での n-gram インデクスプラグインの名称は、_phngram です。

(2) 文字列検索プラグイン

文字列検索プラグインは、FREEWORD プラグインと IXFREEWORD プラグインで構成されています。

FREEWORD プラグイン

FREEWORD プラグインは、データ型を追加するプラグインです。このプラグインは、SQL 抽象データ型を前提としており、FREEWORD 型という抽象データ型を追加して、文字列の検索、登録などの操作機能を実現します。

Text Search Plug-in での FREEWORD プラグインの名称は、_phcfwd です。

IXFREEWORD プラグイン

IXFREEWORD プラグインは、FREEWORD 型に対する検索インデクスを追加するプラグインで、IXFREEWORD 型というインデクス型を追加します。IXFREEWORD プラグインは n-gram インデクス方式の検索エンジンを持ち、FREEWORD プラグインに対して検索手段を提供します。

Text Search Plug-in での IXFREEWORD プラグインの名称は、_phxfwrд です。

さらに、Text Search Plug-in では、次の関連プログラムを提供しています。

Text Search Plug-in Conceptual Extension

Text Search Plug-in Conceptual Extension は、Text Search Plug-in の文書検索プラグインに追加して使用するオプション機能です。Text Search Plug-in Conceptual Extension を組み込むことで、Text Search Plug-in の文書プラグインの構成を変更することなく、概念検索機能を使用できるようになります。

なお、文字列検索プラグインでは概念検索機能は使用できません。

Text Search Enhancer for English

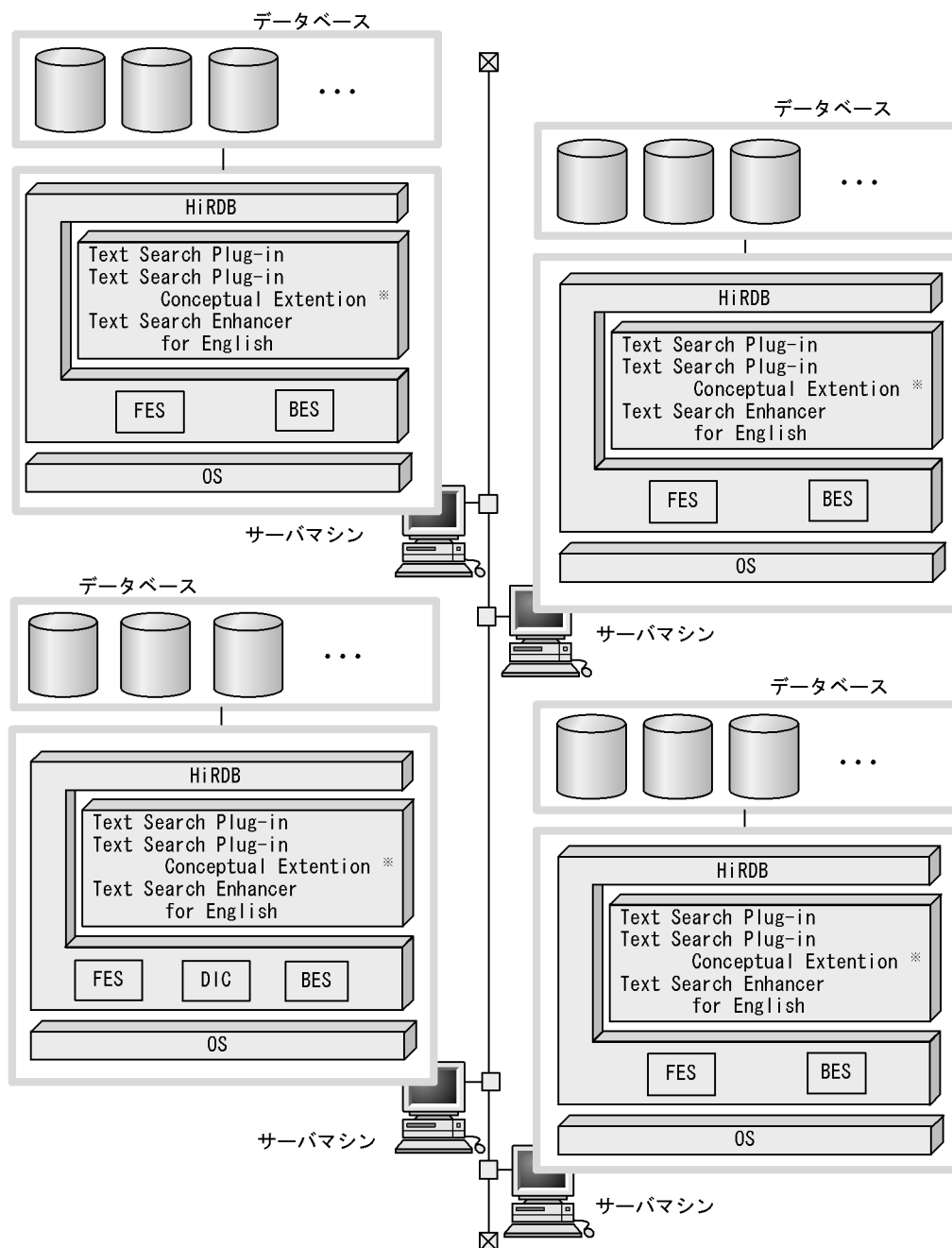
Text Search Enhancer for English は、Text Search Plug-in の文書検索プラグインに追加して使用するオプション機能です。Text Search Enhancer for English を組み込むことで、英単語異表記展開を含む検索の実行を可能にし、もれの無い英文検索機能を提供します。

1.2.2 Text Search Plug-in システムの構成例

Text Search Plug-in システムの構成例を次の図に示します。なお、この例では、HiRDB/ パラレルサーバを前提に説明しています。

1. 解説

図 1-1 Text Search Plug-in システムの構成例



注 インナレプリカ機能を使用する場合は、図中のシステム構成に加え、Staticizer Optionが必要です。また、XML文書を検索する場合には、Preprocessing Library for Text Search Version 2 02-01以降が必要です。

注※ 概念検索機能を使用する場合に必要なプログラムです。

1.3 Text Search Plug-in を利用した検索

Text Search Plug-in を利用した検索には、「全文検索」と、Text Search Plug-in Conceptual Extension を組み込むことで使用できる「概念検索」とがあります。この節では、Text Search Plug-in での検索の概要について説明します。

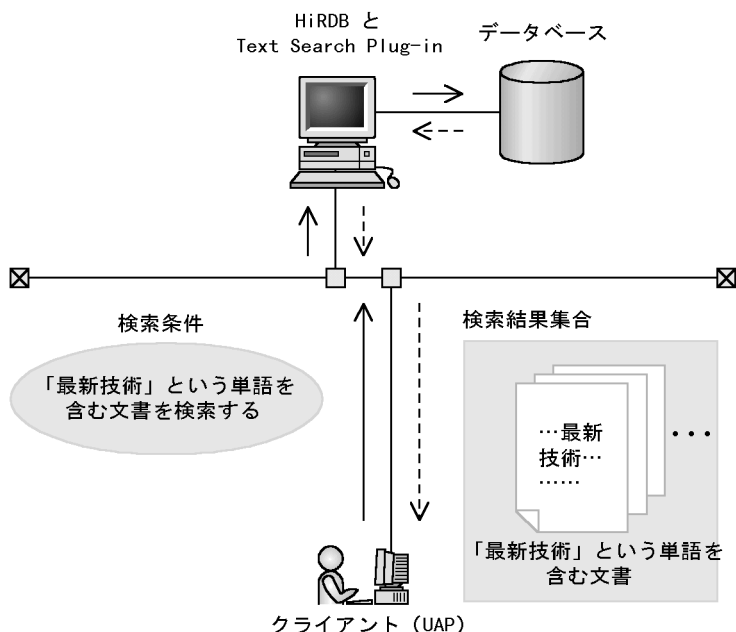
1.3.1 全文検索

Text Search Plug-in の文書検索プラグイン（または文字列検索プラグイン）では、ユーザが任意に指定した単語をキーワードにして文書（または文字列データ）を検索します。この検索方法を全文検索といいます。例えば、キーワードを「最新技術」として、「『最新技術』という単語を含む文書（または文字列データ）を探す」という検索ができます。

検索のキーワードになる単語を、検索タームといいます。検索タームにさらに条件を付け加えることで、さまざまな条件での検索ができます。

Text Search Plug-in での全文検索の概要を次の図に示します。

図 1-2 Text Search Plug-in での全文検索の概要



(凡例)

—> : 検索を要求する制御の流れ

---> : 検索結果を返す制御の流れ

(1) 全文検索の特長

Text Search Plug-in の文書検索プラグインでは、文書を HiRDB の表の一つの列に格納して管理しています。このため、文書の関連情報を、文書を格納した列とは別の列に格納して管理できます。文書を検索する場合に、全文検索条件に加えて、これらの関連情報の値の検索条件を組み合わせることで指定できます。また、文書の関連情報（文字列データ）を文字列検索プラグインに格納することもできます。

(2) 全文検索の種類

Text Search Plug-in での全文検索の基本は、検索タームを指定した検索方法です。この検索方法を基に Text Search Plug-in で提供しているさまざまな検索機能を利用することで、文書または文字列データを柔軟に検索できます。Text Search Plug-in で提供する検索条件の種類を次に説明します。

(a) 検索タームを複数指定する検索

検索条件には、検索タームを複数個指定できます。例えば、「『最先端技術』および『パソコン』の両方の単語が含まれる文書（または文字列データ）」を検索できます。この例では、検索ターム同士の論理積（AND 条件）を検索の条件にしています。さらに、論理和（OR 条件）を検索条件にして、「『最先端技術』または『パソコン』のうち、どちらかを含む文書（または文字列データ）」を検索できます。

(b) 近傍条件検索

二つの検索ターム間の文字数（距離）を検索条件にできます。例えば、「『最新』と『技術』」の間の文字数が 20 文字ちょうどの文字列を含む文書（または文字列データ）」を検索できます。このような検索を近傍条件検索といいます。

(c) 同義語・異表記展開検索

例えば、「s k i」を検索タームにする場合、「ski」「S k i」などで表記されている文書（または文字列データ）も検索したい場合があります。同義語・異表記展開検索はこのような場合に使用します。

同義語展開検索では、同義語辞書の定義に基づいて、検索タームの同義語が自動的に展開されます。異表記展開検索では、Text Search Plug-in でのルールに基づいて自動的に検索タームの異表記が展開されます。展開された同義語・異表記は、検索タームとあわせて検索条件にできます。

HiRDB の LIKE 述語には、同義語・異表記展開の機能はありません。文字列データの同義語・異表記展開検索を実行する場合は、文字列検索プラグインを使用してください。

(d) 論理演算検索

検索条件式同士の論理演算を条件にする検索を、論理演算検索といいます。例えば、「『条件 A』および『条件 B』」で、両方の条件に合致する文書（または文字列データ）」を検索できます。この例では、検索条件同士の論理積（AND 条件）を検索条件にしていま

す。さらに、論理和 (OR 条件) を検索の条件にして、「『条件 A』または『条件 B』のうち、どちらかの条件に合致する文書 (または文字列データ)」を検索できます。

(e) 構造名を指定した検索

文書検索プラグインで構造化文書を格納した表の列を検索対象とする場合は、その表の構造化文書を格納した列に定義された文書の構造を検索条件として利用できます。例えば、「文書・章・節・項」という構造を持つ文書を格納する列があるとします。この列に対しては、「節の中に『SGML』という単語を含む文書」のように、構造を条件にして検索できます。なお、文字列検索プラグインでは検索条件に構造を指定できません。

(f) 重み付き検索

検索条件に複数のキーワードを指定する場合、それぞれの条件に対して「重み」(重要度)を指定できます。

例えば、「政治」と「経済」という二つのキーワードが含まれる文書を検索する場合に、「経済」というキーワードより、「政治」というキーワードの重要度を上げて検索できます。この場合、「政治」の出現回数によって得点 (スコア) を付けた検索結果が得られます。なお、文字列検索プラグインでは検索条件に重みを指定できません。

(g) NOT 検索

検索タームを含まない文書 (または文字列データ) を検索できます。例えば、「『イントラネット』を含まない文書 (または文字列データ)」を検索できます。このような検索を、NOT 検索といいます。ただし、スコア情報は取得できません。

(h) ほかの列の検索条件との複合条件検索

検索タームを指定した条件に加えて、文書以外の情報を格納した列に対する検索条件を組み合わせると検索できます。例えば、文書 (または文字列データ) を格納している列と文書のページ数を格納している列を持つ表に対して、「『先端技術』という単語を含み、ページ数が 200 以下である文書 (または文字列データ)」を検索できます。

(i) 除外文字検索

検索タームの一部または前後に特定の文字がある場合を除外して、該当する文書を検索できます。例えば、「ローマ」を含む文書の中で、「ローマ字」だけを含む文書を除外して検索できます。

1.3.2 概念検索

Text Search Plug-in の文書検索プラグインでは、関連プログラムとして提供している Text Search Plug-in Conceptual Extension を組み込むことで、ユーザが任意に指定した文章や文字列を手がかりにして、その条件と似た概念を持つ文書を検索します。この検索方法概念検索といいます。例えば、「『近年、環境保護に関する対策に力を入れている自治体が増えている』という概念を持つ文書を探す」などの検索ができます。ただし、文字列検索プラグインでは、概念検索機能は使用できません。

1. 解説

概念検索の検索条件に指定する文章を、種文章といいます。概念検索では、初めに種文章を特徴付ける単語が、種文章から抽出されます。この単語を特徴タームといいます。次に、抽出された特徴タームの中から、種文章の概念を表す（実際の検索に使用する）タームが選出されます。ここで選出された特徴タームを検索用特徴タームといいます。

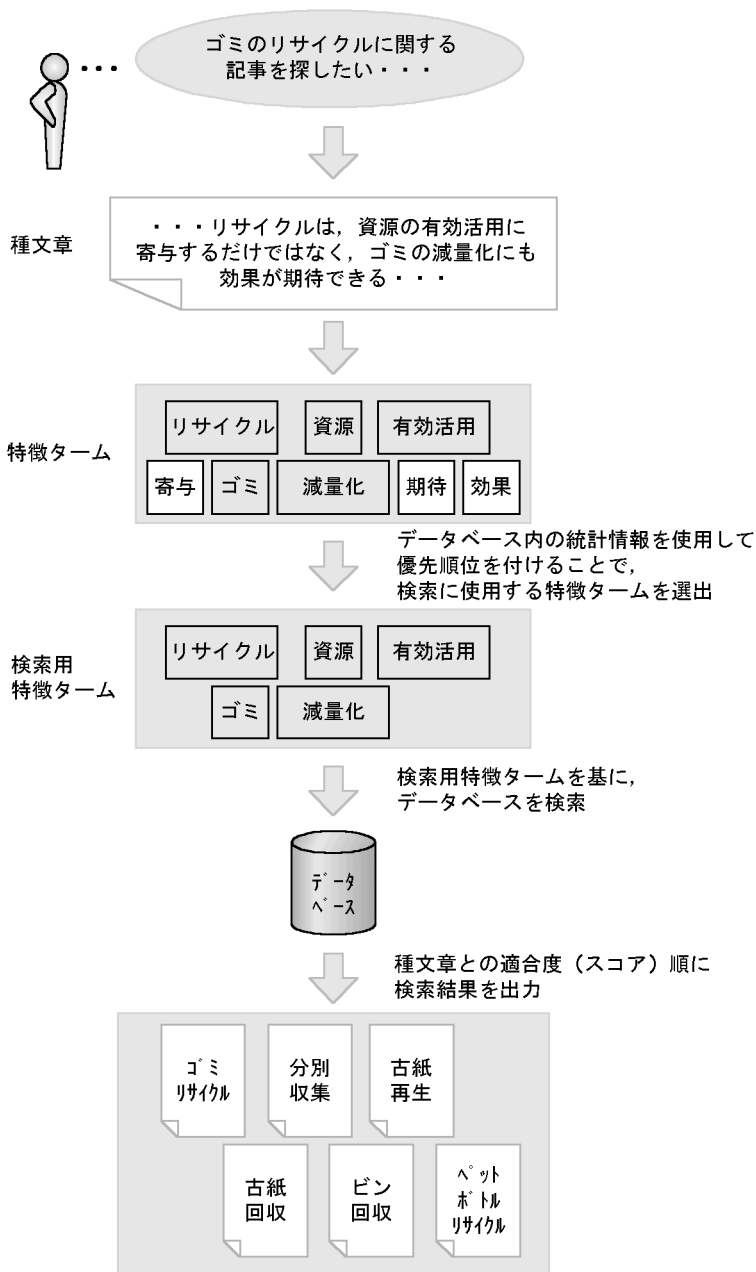
検索用特徴タームの抽出には、データベース内の統計情報を使用します。検索用特徴タームは、次の優先順位に従って、環境定義ファイルの「検索用特徴ターム数」に指定された個数分を上限として、種文章から抽出されます。

- 種文章に多く出現する特徴タームほど優先順位が高くなる
- データベース中に存在する数が少ない特徴タームほど優先順位が高くなる

上記の優先順位に従って抽出された検索用特徴タームを基に、データベース内の文書が検索されます。

Text Search Plug-in Conceptual Extension を組み込んだ場合の概念検索の概要を次の図に示します。

図 1-3 Text Search Plug-in Conceptual Extension を組み込んだ場合の概念検索の概要



(1) 概念検索の特長

概念検索には、次の3種類の特長があります。

1. 解説

(a) 文書検索の簡易化

概念検索では、探したい情報に関係の深い文章（種文章）を検索条件として指定します。複雑な検索条件式を指定する必要はありません。そのため、検索条件のキーワードが限定できない場合や、複雑な検索条件の入力などの操作に不慣れな場合でも、文章の持つ概念から目的の文書を検索できます。

検索条件となる文章は、ユーザが直接入力したり、データベースの登録文書から切り出して指定したりできます。

(b) 文章指定による幅広い検索

概念検索では、文章に記述された内容（概念）に似た文書を検索します。そのため、キーワード検索よりも幅広く目的の文書を捜し出すことができます。

この特長を生かした適用例としては、特許出願時の類似特許検索などが挙げられます。

(c) 新たな思考や発想の支援

概念検索の検索結果を新たな種文章の素材として利用することもできます。このような手順で検索を繰り返すことによって、新たな思考や発想の手がかりを見つけ出すことができます。

(2) 概念検索の種類

Text Search Plug-in で実行できる概念検索の種類を次に説明します。

(a) 種文章を複数指定する検索

検索条件には、種文章を複数個指定できます。例えば、「近年、環境保護に関する対策に力を入れている自治体が増えている」という文章と「リサイクルは、資源の有効活用に寄与するだけでなく、ゴミの減量化にも効果が期待できる」という文章を種文章として指定し、それぞれの概念の和で表される仮想的な種文章に似た概念を持つ文書を検索できます。

(b) 同義語・異表記展開検索

概念検索での同義語展開検索では、同義語辞書での定義に基づいて、検索用特徴タームの同義語が自動的に展開されます。概念検索での異表記展開検索では、Text Search Plug-in でのルールに基づいて、自動的に検索用特徴タームの異表記が展開されます。展開された同義語・異表記は、検索用特徴タームとあわせて検索条件となります。

(c) 論理演算検索

概念検索での論理演算検索は、例えば、『近年、環境保護に関する対策に力を入れている自治体が増えている』という文章と似た概念を持つ文書を検索した結果と、『リサイクルは、資源の有効活用に寄与するだけでなく、ゴミの減量化にも効果が期待できる』という文章と似た概念を持つ文書を検索した結果の両方に含まれる文書を検索できます。この例では、検索条件同士の論理積（AND 条件）を検索条件にしています。

また、論理和（OR 条件）を検索の条件にして、『近年、環境保護に関する対策に力を入れている自治体が増えている』という文章と似た概念を持つ文書を検索した結果と、『リサイクルは、資源の有効活用に寄与するだけでなく、ゴミの減量化にも効果が期待できる』という文章と似た概念を持つ文書を検索した結果のどちらかに含まれる文書を検索することもできます。

(d) 構造名を指定した検索

構造化文書を格納した表の列を検索対象とする場合は、その表の構造化文書を格納した列に定義された文書の構造を検索条件として利用できます。例えば、「文書・章・節・項」という構造を持つ文書を格納する列があるとします。この列に対しては、「節の中に『近年、環境保護に関する対策に力を入れている自治体が増えている』という文章と似た概念を持つ文書」のように、構造を条件にして検索できます。

(e) スコア検索

概念検索の結果文書に、種文章の概念に対する適合度から算出した得点（スコア）を付けます。適合度が高い文書ほど得点が高くなります。

スコア算出時には、例えば、「種文章を 100 点として、相対的な値に得点を調整する（スコアを正規化する）」、「一定のスコア以上の文書だけを検索結果とする」などの指定もできます。

(f) 検索用特徴ターム出力

検索条件として指定した種文章から抽出される、検索用特徴タームを出力します。出力した検索用特徴タームは、検索用特徴タームを指定して検索する際の検索条件として利用できます。また、検索結果の妥当性を判断する材料としても利用できます。

(g) 検索用特徴タームを指定した検索

種文章から抽出された検索用特徴タームを修正し、それを検索条件として利用できます。検索用特徴ターム出力機能を使用して出力した検索用特徴タームに対して、タームの追加、削除、またはタームごとの重みを変更するなどの修正を加え、この修正した検索用特徴タームを条件に検索できます。

(3) 概念検索と全文検索の長所を利用した検索

検索タームによる絞り込み検索

概念検索で文書を抽出しておき、その抽出結果に対して検索タームで絞り込むという、概念検索と全文検索の長所を利用した検索を実現できます。この機能を利用することで、概念検索結果が散漫になることを防止できます。

1.4 Text Search Plug-in で効果的に検索するには

この節では、Text Search Plug-in でより効果的な検索をするための機能について説明します。これらの機能は、インデクス定義で利用の有無を設定します。設定方法については、文書検索プラグインの場合は「2.5.6 インデクスの定義」を、文字列検索プラグインの場合は「2.7.4 インデクスの定義」を参照してください。

(1) 文字種の組み合わせの追加

情報検索時には、例えば、「WWW サーバ」のようなアルファベットやかたかななど、さまざまな種類の文字（文字種）を組み合わせた検索タームを指定する場合があります。このような場合、インデクス作成ルール追加機能で文字種の組み合わせをインデクス情報に追加しておきます。こうすることで、さまざまな文字種の組み合わせで構成される検索タームを指定したときの検索性が向上します。

特に、任意の文章や文字列を指定する概念検索では、インデクス作成ルール追加機能を使用すると検索精度が向上します。

(2) 大小文字または全角半角文字の表記差の吸収

検索対象となる文書情報には、大文字と小文字、全角文字と半角文字のように、表記の異なる文字が含まれています。この表記の違いによって、期待する検索結果が得られない場合があります。例えば、「BaseBall」と指定して検索すると、「BASEBALL」や「baseball」が検索されなかったりします。この大小文字または全角半角文字表記の違いを吸収した検索結果を得るためには、大小文字・全角文字半角文字統一機能を利用します。

この機能には、次の3種類の利用方法があります。

英大文字と英小文字を区別しない（大小文字統一機能の使用）

全角文字と半角文字の英数字、かたかな、および記号を区別しない（全角文字半角文字統一機能の使用）

英大文字と英小文字を区別しない。かつ全角文字と半角文字の英数字、かたかな、および記号を区別しない（大小文字・全角文字半角文字統一機能の使用）

これらのうち、例えば、大小文字統一機能の使用を設定して「book」と指定して検索すると、「Book」や「BOOK」も検索できるようになります。

(3) 拗音の統一

「あ」と「ぁ」、「や」と「ゃ」など、拗音の違いを吸収して検索できる機能を拗音統一機能といいます。この機能を使用すると、次の拗音が統一の対象となります。

●全角文字

あ, い, う, え, お, つ, や, ゆ, よ, ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

●半角文字

ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

例えば、「りょう」を検索タームに指定した場合、「りょう」、「りょう」、「りょう」、および「りょう」が検索対象となります。

(4) 濁音の統一

「バ」と「ハ[゜]」、「ぱ」と「は[゜]」のように濁音や半濁音の違いを吸収して検索できる機能を濁音統一機能といいます。

(5) 検索に使用しない文字の削除

検索に使用しない文字を設定する機能を、削除文字指定機能といいます。例えば、「XML」のように、文字間に半角スペースを入れて表記されている場合、検索タームに半角スペースを入れなければヒットしません。

削除文字指定機能を利用すると、次の文字が登録した文書情報から削除されます。

全角文字の "。(句点), "。(読点), 全角スペース,

半角文字の "。(句点), "。(読点), 半角スペース (0x20), 改行コード (0x0a, 0x0d), タブコード (0x09)

これによって、「XML」のように、検索タームにスペースを入れなくてもヒットさせることができます。

2

Text Search Plug-in システムの構築

この章では、Text Search Plug-in の運用を開始するために必要な定義ファイルの概要、Text Search Plug-in の環境設定およびデータベースの作成について説明します。また、ユーザごとの環境設定方法についても説明します。

-
- 2.1 Text Search Plug-in システムの構築の流れ

 - 2.2 システムの設計

 - 2.3 リソースの見積もり

 - 2.4 Text Search Plug-in の環境設定

 - 2.5 文書検索プラグインのデータベースの作成

 - 2.6 文書検索プラグインのユーザ環境設定

 - 2.7 文字列検索プラグインのデータベースの作成

 - 2.8 文字列検索プラグインのユーザ環境設定

 - 2.9 システムの制限

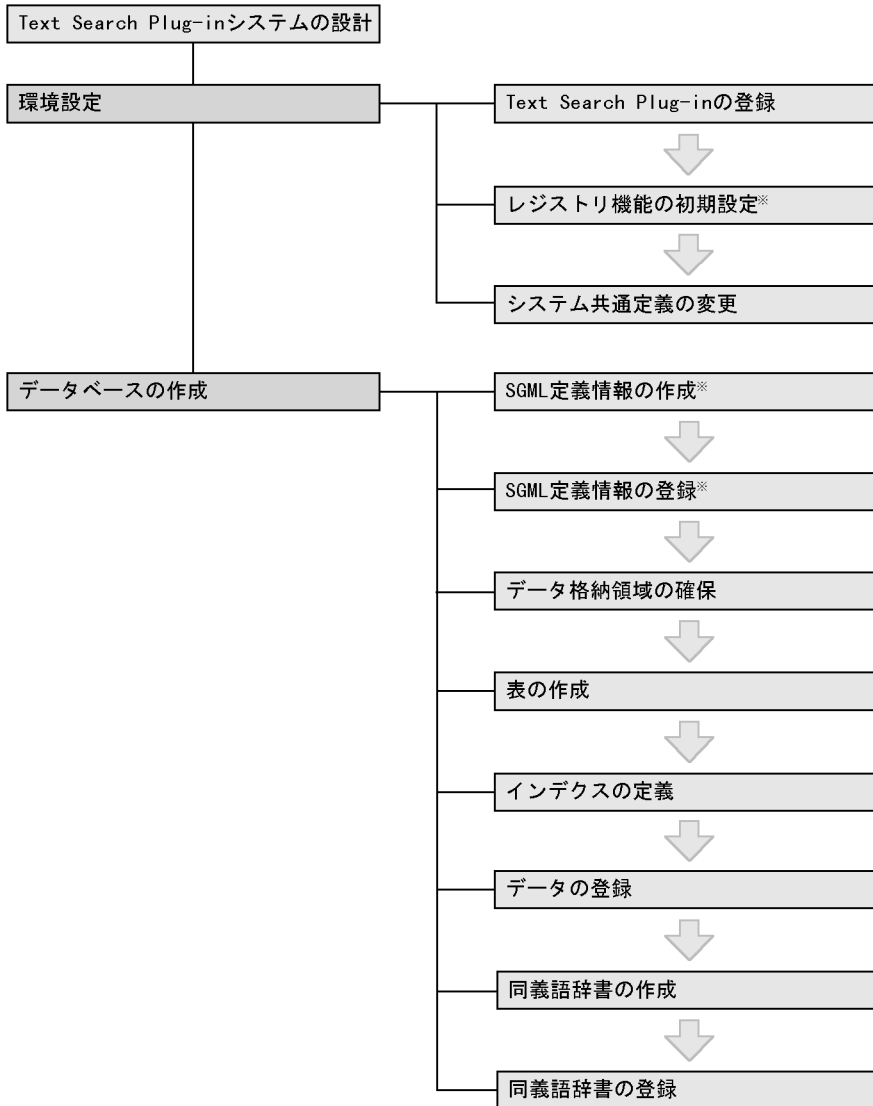
 - 2.10 Text Search Plug-in の設定解除

 - 2.11 HiRDB の提供する機能の利用
-

2.1 Text Search Plug-in システムの構築の流れ

Text Search Plug-in システムの環境を構築する流れを次の図に示します。図に示した構築の流れに従って、次の節以降でシステム構築の方法を説明します。

図 2-1 Text Search Plug-in システムの構築の流れ



注※ 文字列検索プラグインを使用する場合には必要ありません。

注意

Text Search Plug-in を運用するすべてのサーバでは、必ず次の情報を設定してください。

- システム単位，ユニット単位，BES 単位（またはシングルサーバ単位）に設定する環境定義情報
- 同義語展開検索をする場合に使用する同義語辞書情報

これらの情報を設定していないと、検索結果に誤りが生じたり、検索が実行できなかったりする場合がありますのでご注意ください。

2.2 システムの設計

Text Search Plug-in のシステムを構築する前に、検討していただきたいことについて説明します。

2.2.1 検討項目

(1) リソースの見積もり

メモリ所要量や RD エリア容量など、Text Search Plug-in が使用するリソースを見積もってください。リソースの見積もりの詳細については、「2.3 リソースの見積もり」を参照してください。

(2) 検索機能の定義

全文検索または概念検索実行時に次の機能を利用する場合、インデクスの定義時に機能使用の設定が必要です。用途に合わせて設定してください。

インデクス作成ルール追加機能

英字やかたかななど、さまざまな種類の文字種の組み合わせをインデクス情報に追加する機能です。インデクス作成ルール追加機能を使用した場合の長所および短所は次のとおりです。

- 長所

インデクス作成ルール追加機能を使用しない場合と比べて、検索タームがさまざまな文字種の組み合わせで構成される場合の全文検索性能が向上します。また、種文章から抽出された検索用特徴タームがさまざまな文字種の組み合わせで構成される場合の概念検索性能および精度が向上します。

- 短所

インデクスサイズが増加します。

大小文字・全角文字半角文字統一機能

大小文字または全角半角文字表記の違いを吸収した検索結果を得るための機能です。大小文字、全角文字半角文字を統一するので、異表記展開の全角半角異表記展開、および半角アルファベット展開を実行する必要がありません。異表記展開を実行した場合と比較して、検索性能が向上します。

拗音統一機能

「ゃ」と「や」、「ょ」と「よ」など、拗音による違いを吸収した検索結果を得るための機能です。拗音を統一するので、異表記展開のひらがな拗音展開を実行する必要はありません。異表記展開を実行した場合と比較して、検索性能が向上します。

濁音統一機能

「バ」や「ば」などの濁音を表記する方法として「ハ`」や「は`」のように濁音や半濁音を分離して表記する方法もあります。濁音統一機能を使用することにより、このような表記の違いを吸収して検索することが可能です。

削除文字指定機能

検索に使用しない文字を設定する機能です。例えば、「X M L」のように、文字間に半角スペースを入れて表記されている場合、検索タームに半角スペースを入れなければヒットしません。しかし、削除文字指定機能を利用すると、スペースなどが登録した文書情報から削除されるため、検索タームにスペースを入れなくてもヒットさせることができます。

また、この機能を利用すると、インデクス容量を削減できます。

(3) 運用形態の決定

Text Search Plug-in の運用方法を検討します。推奨する運用方法については「5. 推奨する運用方法」を参照してください。

インデクス定義時に、次の登録運用の形態を設定できます。用途に合わせて設定してください。

(a) 差分インデクス - 登録性能の向上

小容量の一時的な登録用インデクスに、文書を追加する機能を提供します。この登録用インデクスを差分インデクスといいます。

n-gram インデクスプラグインは、登録しているインデクス量が少ないほど、登録性能が良くなります。差分インデクスは、この特性を利用し、小容量の一時的なインデクスに登録することで登録性能を向上させる機能です。これは UAP によるデータ登録をする運用の場合に有効です。差分インデクスを使用しない通常の登録よりも短時間で更新できます。

なお、差分インデクスを定義している場合は、差分インデクスから MASTER インデクスへのマージ処理が必要です。

(b) インデクス作成方法の指定

表の SGMLTEXT 型列に文書を登録している、または FREEWORD 型列に文字列を登録している状態で CREATE INDEX を実行する場合、次のどれかのインデクス作成方法を指定できます。

即時登録

インデクス作成とインデクスへの文書登録を同時に実行する方法です。

初期分割登録

インデクス作成時には文書登録をしないで、インデクス作成開始ユティリティ (phnstartidx または phxstartidx) で実行時間やインデクスへの登録件数を指定し、文書を登録する方法です。初期分割登録の対象となるのは、インデクス作成時に登録済みのデータです。例えば、表に大量の文書を登録している状態でインデクス作成を実行する場合は、インデクスへの文書登録に時間が掛かります。このような場合、時間や件数で区切って登録できます。なお、初期分割登録ではインデクスへの文書登録が完了すると、以降は即時登録となります。

2. Text Search Plug-in システムの構築

分割遅延登録

初期分割登録での対象範囲に加えて、インデクス作成後に登録された文書も遅延登録する方法です。インデクス作成後は初期分割登録と同様に、「インデクス作成開始ユティリティ (phnstartidx または phxstartidx)」を実行して、文書をインデクスへ一括登録できます。また、「インデクス分割遅延登録運用変更ユティリティ (phndelay または phxdelay)」を利用して、分割遅延登録の開始と終了を指定することもできます。

2.2.2 使用する文字コードによる機能差異について

HiRDB が文字コード utf-8 でセットアップされた場合、Text Search Plug-in でも文字コード utf-8 を使用できるようになります。この場合、文字コードが utf-8 である XML 文書、プレーン文書、および FREEWORD 型テキストが検索対象となります。ただし、FREEWORD 型テキストについては、文字コード種別が MS-Unicode である必要があります。(MS-Unicode は Microsoft 社の Unicode です。)

また、使用する文字コードによって、Text Search Plug-in で使用できる機能が異なります。使用する文字コードによる Text Search Plug-in の機能差を、表 2-1、表 2-2 および表 2-3 に示します。

表 2-1 使用できる検索機能

| 機能 | Shift-JIS または EUC 使用時 | utf-8 使用時 |
|------------------------------|-----------------------|-----------|
| 単純文字列検索 | | |
| 検索タームを複数指定した検索 (論理和または論理積使用) | | |
| 近傍検索 | | |
| 同義語展開検索 | | |
| 異表記展開検索 | | |
| 構造名を指定した検索 | | |
| 重み付き検索 | | |
| 概念検索 | | |
| 抽象データ型関数 extracts | | 1 |

(凡例)

: 使用できる。

: 制限付きで使用できる。

注

HiRDB が文字コード utf-8 の場合、TSPlugin の下で使用できるパス名、ファイル名は ASCII の範囲内の必要があります。

注 1

文字コード utf-8 で extracts 関数を使用する場合、XML 出力を利用してください (SGML 出力はできません)。また、extracts 関数に渡す引数は必ず MS-Unicode で記述してください。

表 2-2 インデクス作成時に使用できる機能

| 機能 | Shift-JIS または EUC 使用時 | utf-8 (UCS2) 使用時 | utf-8 (UCS4) 使用時 |
|-------------------|-----------------------|------------------|------------------|
| 概念検索の使用 | | | |
| インデクス作成ルール追加機能 | | | |
| 大小文字・全角文字半角文字統一機能 | | | |
| 拗音統一機能 | | | |
| 濁音統一機能 | | | |
| 削除文字指定機能 | | | |
| 文字単位インデクス作成機能 | | | |
| 差分インデクス定義 | | | x |
| 分割遅延登録時のインデクス定義 | | | |
| 英文インデクス定義 | | | x |

(凡例)

: 使用できる。

x : 使用できない。

表 2-3 使用できるユティリティ

| プラグイン種別 | ユティリティ コマンド名 | 機能 | Shift-JIS または EUC 使用時 | utf-8 (UCS2) 使用時 | utf-8 (UCS4) 使用時 |
|-------------------|--------------------------|--------------|-----------------------------|------------------------|------------------------|
| 文書検索 プラグ イン | phsregsetup | レジストリの初期化 | | | |
| | phssgmlreg | SGML 定義情報の登録 | | | |
| | phssgmlrm | SGML 定義情報の削除 | | | |
| | phssgmlval | SGML 定義情報の参照 | | | |
| | phssyndicreg | 同義語辞書の登録 | | | |
| | phssyndicrm | 同義語辞書の削除 | | | |
| | phssyndiconv | 同義語辞書の移行 | | x | x |
| | phssyndicbkup (UNIX の場合) | 同義語辞書のバックアップ | | | |

2. Text Search Plug-in システムの構築

| プラグイン種別 | ユーティリティ コマンド名 | 機能 | Shift-JIS または EUC 使用時 | utf-8 (UCS2) 使用時 | utf-8 (UCS4) 使用時 | |
|----------------------------------|---------------------------|-------------------------------------|--------------------------------|--------------------------|--------------------------|---|
| n-gram インデ クスプ ラグイ ン | phnidxls | インデクス情報の 取得 | | | | |
| | phnincmt | インデクスのイン クリメンタル (最 適化) | | | × | |
| | phnmerge | 差分インデクスか ら MASTER イン デクスへのマージ | | | × | |
| | phnexpnd | インデクスの拡張 | | | × | |
| | phnstartidx | インデクスの作成 開始 | | | | |
| | phndelay | インデクス遅延更 新開始 / 終了指定 | | | | |
| | phnmodidx | インデクス情報変 更 | | | | |
| | phncond | インデクス詰め替 え | | | | |
| | phnconget | 統計情報収集 | | × | × | |
| | phnconbuild | 統計情報統合 | | × | × | |
| | phnconrstr | 統計情報反映 | | × | × | |
| | phnchgcon | 特徴ターム取得 RD エリア指定 | | | | |
| 文字列 検索プ ラグイ ン | FREE WORD プラグ イン | phcsyndicreg | 同義語辞書の登録 | | | |
| | | phcsyndicrm | 同義語辞書の削除 | | | |
| | | phcsyndicconv | 同義語辞書の移行 | | × | × |
| | | phcsyndicbkup (UNIX の 場合) | 同義語辞書のバッ クアップ | | | |
| IXFRE E WORD プラグ イン | | phxidxls | インデクス情報の 取得 | | | |
| | | phxincmt | インデクスのイン クリメンタル (最 適化) | | | × |

| プラグイン種別 | ユーティリティ コマンド名 | 機能 | Shift-JIS または EUC 使用時 | utf-8 (UCS2) 使用時 | utf-8 (UCS4) 使用時 |
|---------|------------------|------------------------|-----------------------------|------------------------|------------------------|
| | phxexpnd | インデクスの拡張 | | | × |
| | phxstartidx | インデクスの作成 開始 | | | |
| | phxdelay | インデクス遅延更 新開始 / 終了指定 | | | |
| | phxmodidx | インデクス情報変 更 | | | |
| | phxcond | インデクス詰め替 え | | | |

(凡例)

: 使用できる。

× : 使用できない。

注

文字コード utf-8 (UCS4) 使用時に次のユーティリティを実行した場合、処理を行わずに正常終了します。

- phnmerge (差分インデクスから MASTER インデクスへのマージ)
- phnexpnd, phxexpnd (インデクスの拡張)

2.3 リソースの見積もり

ここでは、Text Search Plug-in が使用するリソースの所要量の見積もりについて説明します。

2.3.1 メモリ所要量

Text Search Plug-in のメモリ所要量の見積もり方法について説明します。

(1) SGML プラグインと n-gram インデクスプラグインの場合

(a) データ操作 (文書登録 / 更新)

単一文書登録時

遅延更新機能またはインデクス未定義の場合

使用メモリ [メガバイト] = fesprocess + besprocess

インデクス作成の場合

使用メモリ [メガバイト] = $\max(\min(\text{config 指定メモリ} \times 10^{-3}, \text{docsize} \times 5 \times 10^{-6}), \text{fesprocess} + \text{besprocess})$

- $\text{fesprocess} = ((\text{dtdsize}/8192 \times 8192 + \text{norparamsize}/8192 \times 8192) \times 2 + 8192) \times 10^{-6}$
- $\text{besprocess} = ((\text{dtdsize}/8192 \times 8192 + \text{norparamsize}/8192 \times 8192) + \text{docsize} \times 5) \times 10^{-6}$
- config 指定メモリ: n-gram インデクスプラグインの環境設定項目「1件登録時最大メモリ」の設定値 [キロバイト]
- fesprocess: FES での使用メモリ [メガバイト]
- besprocess: BES での使用メモリ [メガバイト]
- dtdsize: DTD サイズ [バイト]
- norparamsize: 正規化パラメタサイズ [バイト]
- docsize: 登録文書サイズ [バイト]

複数文書一括登録

差分インデクス統合, CREATE INDEX, 遅延更新時のインデクス作成の場合

使用メモリ [メガバイト] = $\min(\text{config 指定メモリ} \times 10^{-3}, \text{docsize} \times \text{doccnt} \times 5 \times 10^{-6})$

pload の場合

使用メモリ [メガバイト] = $\max(\text{docsize} \times 5 \times \text{SGMLTEXT 型定義列数} \times 10^{-6}, \min(\text{config 指定メモリ} \times 10^{-3}, \text{docsize} \times \text{doccnt} \times 5 \times 10^{-6}))$

- docsize: 平均登録文書サイズ [バイト]
- doccnt: 登録文書数

- config 指定メモリ：n-gram インデクスプラグインの環境設定項目「一括登録時最大メモリ」の設定値 [キロバイト]
- (b) データ操作 (文書削除)
- 使用メモリ [キロバイト] = 24
- (c) データ操作 (インデクス情報の取得ユティリティ)
- 使用メモリ [メガバイト] = 2
- (d) データ操作 (インデクスのインクリメンタルユティリティ)
- 使用メモリ [メガバイト] = config 指定メモリ $\times 10^{-3}$
- config 指定メモリ：n-gram インデクスプラグインの環境設定項目「一括登録時最大メモリ」の設定値 [キロバイト]
- (e) データ操作 (インデクス詰め替えユティリティ)
- 使用メモリ [メガバイト] = 10
- (f) データ操作 (統計情報収集ユティリティ)
- 収集モード指定：normal
- 使用メモリ [メガバイト] = 5
- 収集モード指定：all
- 使用メモリ [メガバイト] = config 指定メモリ $\times 10^{-3}$
- config 指定メモリ：n-gram インデクスプラグインの環境設定項目「一括登録時最大メモリ」の設定値 [キロバイト]
- (g) データ操作 (統計情報統合ユティリティ)
- 使用メモリ [メガバイト] = 20
- (h) データ操作 (統計情報反映ユティリティ)
- 使用メモリ [メガバイト] = 10
- (i) 検索
- 差分インデクスを使用していない場合
- 1 回の全文検索での使用メモリ [メガバイト] = $\min (\text{config 指定メモリ} \times 10^{-3}, \text{ヒット件数} \times \text{検索ターム数} \times \text{検索ターム平均長} / 2 \times 20 \times 10^{-6}) + \text{ヒット件数} \times 12 \times 10^{-6} + \text{同義語異表記展開用メモリ} + 3$
 - 1 回の概念検索での使用メモリ [メガバイト] = $\max (\text{種文章数} \times \text{種文章の平均長} [\text{バイト}] \times 8 \times 10^{-6}, \min (\text{config 指定メモリ} \times 10^{-3}, \text{ヒット件数} \times \text{config 検索用特徴ターム数} \times 60 \times 10^{-6})) + \text{ヒット件数} \times 12 \times 10^{-6} + \text{同義語異表記展開用}$

2. Text Search Plug-in システムの構築

メモリ + 3

差分インデクスを使用している場合

- 1回の全文検索での使用メモリ [メガバイト] = $\min(\text{config 指定メモリ} \times 10^{-3}, \text{ヒット件数} \times \text{検索ターム数} \times \text{検索ターム平均長} / 2 \times 20 \times 10^{-6}) + \text{ヒット件数} \times 12 \times 10^{-6} + \text{同義語異表記展開用メモリ} + 5$
- 1回の概念検索での使用メモリ [メガバイト] = $\max(\text{種文章数} \times \text{種文章の平均長} [\text{バイト}] \times 8 \times 10^{-6}, \min(\text{config 指定メモリ} \times 10^{-3}, \text{ヒット件数} \times \text{config 検索用特徴ターム数} \times 60 \times 10^{-6})) + \text{ヒット件数} \times 12 \times 10^{-6} + \text{同義語異表記展開用メモリ} + 5$

同義語異表記展開用メモリ

- 全文検索での同義語異表記展開用メモリ [メガバイト] = $((\text{検索ターム数} \times \text{検索ターム平均長} / 2 \times 40 + 24) \times \text{平均同義語異表記展開数} + 24) \times 10^{-6}$
- 概念検索での同義語異表記展開用メモリ [メガバイト] = $((\text{config 検索用特徴ターム数} \times 120 + 24) \times \text{平均同義語異表記展開数} + 24) \times 10^{-6}$
- config 指定メモリ: n-gram インデクスプラグインの環境設定項目「検索時最大ワークメモリサイズ」の設定値 [キロバイト]
- config 検索用特徴ターム数: n-gram インデクスプラグインの環境設定項目「検索用特徴ターム数」

注 平均同義語異表記展開数の見積もりについては、「4.3.6 異表記展開指定」を参照してください。

1SQL でプラグイン関数 (contains 関数, contains_with_score 関数, および concept_with_score 関数) を複数指定した場合

- 検索時のメモリ所要量の最大値「1回の検索使用メモリ × 1SQL 文中に記述したプラグイン関数の個数」

(j) 文書抽出

反転位置情報取得をしない場合

使用メモリ [メガバイト] = $\text{docsize} \times 3 \times 10^{-6}$

反転位置情報取得をする場合

使用メモリ [メガバイト] = $\text{docsize} \times 8 \times 10^{-6}$

- docsize: 登録文書サイズ [バイト]

(k) 1RD エリア当たりの排他制御用共用メモリ

登録・検索同時実行を実行する場合、BES またはシングルサーバ内の n-gram インデクス用の RD エリア単位に、次のサイズの排他制御用共用メモリが必要です。この領域は、n-gram インデクス用の共用メモリ中に確保されます。

使用メモリ [キロバイト] = $\max((\text{user_cnt} \times \text{lock_obj_size} \times 10) \times 10^{-3}, 8)$

user_cnt : HiRDB システムで同時に実行できるユーザ数の最大数 (HiRDB 環境定義 pd_max_users 指定値)

- lock_obj_size : 1 排他オブジェクト当たりのサイズ [バイト]

UNIX の場合

64 ビット版 Text Search Plug-in の場合 : 48

64 ビット版以外の Text Search Plug-in の場合 : 32

Windows の場合

32 バイト固定

(1) 共用メモリ

登録・検索同時実行を実行する場合, BES またはシングルサーバ単位に, 次のサイズの共用メモリが必要です。

使用メモリ [キロバイト] = $\max((\text{rd_serve} \times 8192 \times \text{lock_size} / 8) / 1024 + (32+16 \times \text{rd_serve}) / 8192 \times 8, 16)$

- rd_serve : 1 サーバ内の n-gram インデクス用 RD エリア数
- lock_size : n-gram インデクスプラグインの環境設定項目「1RD エリア当たりの排他制御用共用メモリサイズ」の設定値 [キロバイト]

(2) FREEWORD プラグインと IXFREEWORD プラグインの場合

(a) データ操作 (文字列データ登録 / 更新)

単一文字列データ登録時

- 遅延更新機能またはインデクス未定義の場合
使用メモリ [メガバイト] = fesprocess + besprocess
- インデクス作成の場合
使用メモリ [メガバイト] = $\max(\min(\text{config 指定メモリ} \times 10^{-3}, \text{freewordsize} \times 5 \times 10^{-6}), \text{fesprocess} + \text{besprocess})$
- fesprocess = 0
- besprocess = $(\text{docsize} \times 3) \times 10^{-6}$
- config 指定メモリ : IXFREEWORD プラグインの環境設定項目「1件登録時最大メモリ」の設定値 [キロバイト]
- fesprocess : FES での使用メモリ [メガバイト]
- besprocess : BES での使用メモリ [メガバイト]
- freewordsize : 登録文字列データ長 [バイト]

複数文字列データ一括登録

- CREATE INDEX, 遅延更新時のインデクス作成の場合
使用メモリ [メガバイト] = $\min(\text{config 指定メモリ} \times 10^{-3}, \text{freewordsize} \times$

2. Text Search Plug-in システムの構築

$\text{freewordcnt} \times 5 \times 10^{-6}$)

- pdload の場合

使用メモリ [メガバイト] = $\max(\text{freewordsize} \times 5 \times \text{FREEWORD 型定義列数} \times 10^{-6}, \min(\text{config 指定メモリ} \times 10^{-3}, \text{freewordsize} \times \text{freewordcnt} \times 5 \times 10^{-6}))$

- freewordsize : 平均登録文字列データ長 [バイト]
- freewordcnt : 登録文字列データ件数
- config 指定メモリ : IXFREEWORD プラグインの環境設定項目「一括登録時最大メモリ」の設定値 [キロバイト]

(b) データ操作 (文字列データ削除)

使用メモリ [キロバイト] = 24

(c) データ操作 (インデクス情報の取得ユティリティ)

使用メモリ [メガバイト] = 2

(d) データ操作 (インデクスのインクリメンタルユティリティ)

使用メモリ [メガバイト] = $\text{config 指定メモリ} \times 10^{-3}$

- config 指定メモリ : IXFREEWORD プラグインの環境設定項目「一括登録時最大メモリ」の設定値 [キロバイト]

(e) データ操作 (インデクス詰め替えユティリティ)

使用メモリ [メガバイト] = 10

(f) 検索

1 回の全文検索での使用メモリ [メガバイト] = $\min(\text{config 指定メモリ} \times 10^{-3}, \text{ヒット件数} \times \text{検索ターム数} \times \text{検索ターム平均長} / 2 \times 20 \times 10^{-6}) + \text{ヒット件数} \times 12 \times 10^{-6} + \text{同義語異表記展開用メモリ} + 3$

同義語異表記展開用メモリ [メガバイト] = $((\text{検索ターム数} \times \text{検索ターム平均長} / 2 \times 40 + 24) \times \text{平均同義語異表記展開数} + 24) \times 10^{-6}$

- config 指定メモリ : IXFREEWORD プラグインの環境設定項目「検索時最大ワークメモリサイズ」の設定値 [キロバイト]

注

平均同義語異表記展開数の見積もりについては、「4.4.4 異表記展開指定」を参照してください。

1SQL でプラグイン関数 (contains 関数) を複数指定した場合

検索時のメモリ所要量の最大値「1 回の検索使用メモリ × 1SQL 文中に記述したプラグイン関数の個数」

(g) 文字列データ抽出

使用メモリ [メガバイト] = $\text{freewordsize} \times 3 \times 10^6$

- freewordsize : 登録文字列データ長 [バイト]

(h) 1RD エリア当たりの排他制御用共用メモリ

登録・検索同時実行を実行する場合、BES またはシングルサーバ内の IXFREEWORD インデクス用の RD エリア単位に、次のサイズの排他制御用共用メモリが必要です。この領域は、IXFREEWORD インデクス用の共用メモリ中に確保されます。

使用メモリ [キロバイト] = $\max((\text{user_cnt} \times \text{lock_obj_size} \times 10) \times 10^3, 8)$

user_cnt : HiRDB システムで同時に実行できるユーザ数の最大数 (HiRDB 環境定義 pd_max_users 指定値)

- lock_obj_size : 1 排他オブジェクト当たりのサイズ [バイト]

UNIX の場合

64 ビット版 Text Search Plug-in の場合 : 48

64 ビット版以外の Text Search Plug-in の場合 : 32

Windows の場合

32 バイト固定

(i) 共用メモリ

登録・検索同時実行を実行する場合、BES またはシングルサーバ単位に次のサイズの共用メモリが必要です。

使用メモリ [キロバイト] = $\max((\text{rd_serve} \times 8192 \times \text{lock_size} / 8) / 1024 + (32+16 \times \text{rd_serve}) / 8192 \times 8, 16)$

- rd_serve : 1 サーバ内の IXFREEWORD インデクス用 RD エリア数
- lock_size : IXFREEWORD プラグインの環境設定項目「1RD エリア当たりの排他制御用共用メモリサイズ」の設定値 [キロバイト]

2.3.2 RD エリア容量

Text Search Plug-in を使用する場合は RD エリア容量の見積もり方法について説明します。

見積もりの必要な RD エリア容量は、ユーザ用 RD エリアと、ユーザ LOB 用 RD エリア、データディクショナリ用 RD エリア、レジストリ用 RD エリアおよびレジストリ LOB 用 RD エリアの容量です。

(1) SGML プラグインおよび n-gram インデクスプラグインの場合

(a) ユーザ用 RD エリア容量

SGMLTEXT 型列を定義した表を格納するユーザ用 RD エリアの容量は、次の値で見積もってください。ユーザ用 RD エリアの見積もり式は、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。

| 設定箇所 | 変数 | 設定値 |
|---------|-----|------------|
| 各列のデータ長 | di= | 26 [バイト] |

(b) ユーザ LOB 用 RD エリア容量

HiRDB のユーザ LOB 用 RD エリアの容量の見積もりでは、次に説明する SGMLTEXT データ格納用 RD エリアと n-gram インデクス情報格納用 RD エリアの容量を計算します。

SGMLTEXT データ格納用 RD エリア

次の値を設定して、SGMLTEXT データを格納するユーザ LOB 用 RD エリアの容量を見積もってください。ユーザ LOB 用 RD エリアの容量の見積もり式は、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。

| 設定箇所 | 変数 | 設定値 |
|-----------------|-----|---------------|
| LOB 列の行の総数 | b= | 2 + 登録文書数 |
| 各 BLOB データのデータ長 | Cj= | 次の説明を参照してください |

C1 のサイズ

$$C_1 = 2,629,632$$

C2 のサイズ

$$C_2 = 1,318,912$$

Cj のサイズ

一般の場合

$$C_j = (\text{docsize} \times 2.5) / 8192 \times 8192 + 8192 (j > 2)$$

DocumentBroker を使用している場合

$$C_j = (\text{docsize} \times 2.5) / 8192 \times 8192 + 8192 + \text{docbrotdtdsize} (j > 2)$$

Preprocessing Library for Text Search を使用して XML 文書を登録する場合

$$C_j = (\text{docsize} \times 1.5) / 8192 \times 8192 + 8192 (j > 2)$$

n : 各タグに対する存在構造数

elen : 要素型名称長

docsize : 1 文書当たりのサイズ [バイト]

docbrotdtdsize : DocumentBroker に定義した DTD サイズ [バイト]

n-gram インデクス情報格納用 RD エリア

次の値を設定して、n-gram インデクス情報を格納するユーザ LOB 用 RD エリアの容量を見積もってください。ユーザ LOB 用 RD エリアの容量見積もり式は、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。差分インデクスまたは分割遅延登録を使用する場合は、指定した差分インデクスサイズまたはインデクス分割遅延登録用領域サイズを加算したサイズとなります。また、1つの RD エリアサイズの最大サイズは 64GB です。64GB を超える場合は表分割してください。

| 設定箇所 | 変数 | 設定値 |
|-----------------|-----|---|
| LOB 列の行の総数 | b= | 5 + + 分割遅延登録を使用する場合 =1 分割遅延登録を使用しない場合 =0 統計情報を反映する場合 =1 統計情報を反映しない場合 =0 |
| 各 BLOB データのデータ長 | Cj= | 次の説明を参照してください |

C1 のサイズ

$$C_1 = \max \left(\left(\frac{24 \times \text{doccnt}}{8192} + \frac{4 \times \text{doccnt}}{512 \times 8192} \right) \times 8192 + 8192, \frac{4 \times \text{doccnt}}{1024 \times 8192} \right) \times 8192 + 8192, 128000 + 8192$$

C2 のサイズ

概念検索を使用する場合

$$C_2 = 4875 \times 8192 + 8192$$

概念検索を使用しない場合

$$C_2 = 3750 \times 8192 + 8192$$

C3 のサイズ

通常インデクスの場合または英文検索用インデクス (ENGLISH) の場合

$$C_3 = \max \left(\frac{\text{doccnt} \times \text{docsize} \times 3}{8192} \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

インデクス作成ルール追加機能 (INDEX_ALL) を使用した場合

$$C_3 = \max \left(\frac{\text{doccnt} \times \text{docsize} \times 7}{8192} \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

インデクス作成ルール追加機能 (INDEX_MEDIUM) または英文検索用インデクス (ENGLISH_STANDARD) を使用した場合

2. Text Search Plug-in システムの構築

$$C_3 = \max \left(\left(\text{doccnt} \times \text{docsize} \times 5 \right) / 8192 \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

文字単位インデクス作成機能 (INDEX_GRAM1) を使用した場合

$$C_3 = \max \left(\left(\text{doccnt} \times \text{docsize} \times 2 \right) / 8192 \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

C4 のサイズ

$$C_4 = 160 \times 8192 + 8192$$

C5 のサイズ

$$C_5 = 160 \times 8192 + 8192$$

C6 のサイズ

CREATEMODE=2 の分割遅延登録用インデクスの場合 (DELAY_SIZE 設定値のインデクス分割遅延登録用領域サイズ)

$$C_6 = C_1 \times 2 + 1024 \times 1024$$

ただし、 C_1 中の doccnt は遅延登録文書数です。

CREATEMODE=3 の分割遅延登録・更新系 SQL 同時実行用インデクスの場合 (DELAY_SIZE 設定値の遅延登録時に情報を退避する領域サイズ)

$$C_6 = 16 \times \text{delaydoccnt} + 12 \times 8192$$

delaydoccnt : 遅延登録文書数

C7 のサイズ

統合統計情報を反映した場合 (1 回目)

$$C_7 = \text{反映する統合統計情報ファイルサイズ (バイト)} / 8192 \times 8192 + 8192$$

統合統計情報を反映した場合 (2 回目以降)

$$C_7 = 1 \text{ 回目に反映したサイズ} + \left(\text{反映する統合統計情報ファイルサイズ (バイト)} - 1 \text{ 回目に反映したサイズ} \right) / 4188 \times 4188$$

C8 のサイズ

UCS4 インデクスの場合 (初期作成)

$$C_8 = 50 \times 1024 \times 1024$$

UCS4 インデクスの場合

次のサイズを Available Work Area Size (使用可能ワーク領域サイズ) ¹ に確保してください。

$$C_8 = 200 \times 1024 \times 1024 + \text{doccnt} \times 1 \text{ 文書中の UCS4 文字数} \times 7$$

注 1

- phnidxls (インデクス情報取得ユティリティ) を使って取得したインデクス

情報に表示されます。HiRDB の pdmod コマンドによる RD エリアの拡張を行うと拡張されます。

注 2

- UCS4 で表現できる文字の中で、UCS2 を超える範囲。

0x00010000 ~ 0x7fffffff

doccnt : 登録文書数

docsize : 平均登録文書長 [バイト]

(c) データディクショナリの RD エリア容量

Text Search Plug-in で、SGMLTEXT 型や NGRAM インデクス型を表に適用する場合は、次の値を設定して HiRDB のデータディクショナリ用 RD エリアの容量を見積もってください。データディクショナリ用 RD エリア容量の見積もり式は、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。

表の格納ページ数の設定項目と設定値

| 設定項目 | 設定値 |
|--------------------------|------------------------------------|
| ユーザ定義型の名称長の平均値 | 8 [バイト] |
| プラグインオプション指定長の平均値 | 72+DTD 名称長 + 正規化パラメタ名称長 [バイト] |
| ユーザ定義型の総数 | 1 [個] |
| 1 データ型当たりの属性数の平均値 | 1 [個] |
| データ型識別子の長さの平均値 | 8 [バイト] |
| ユーザ定義型で定義された属性数 | 1 [個] |
| プラグインの総数 | 2 [個] |
| プラグイン名称の長さの平均値 | 8 [バイト] |
| 抽象データ型 / インデクス型名称の長さの平均値 | 8 [バイト] |
| インデクス型の総数 | 1 [個] |
| インデクス型識別子の長さの平均値 | 4 [バイト] |
| 抽象データ型名称の長さの平均値 | 8 [バイト] |
| プラグインインデクスの総数 | 1 [個] |
| インデクス型識別子の長さの平均値 | 4 [バイト] |
| 抽象データ型名称の長さの平均値 | 8 [バイト] |
| 抽象データ型で定義された属性の総数 | 1 [個] |
| プラグインインデクス適用関数名称の長さの平均値 | 10 [バイト] |
| 1 プラグインインデクス当たりの適用関数の数 | 30 [個] |

2. Text Search Plug-in システムの構築

| 設定項目 | 設定値 |
|--------------------------|-----------------------------------|
| LOB 列を定義した表の総数 | ユーザ LOB 列に加えて SGMLTEXT 型列を定義した表の数 |
| 1 プラグイン当たりのパラメタ名称の長さの平均値 | 30 [バイト] |
| プラグインライブラリパス名称の長さの平均値 | 255 [バイト] |
| プラグインの注釈の長さの平均値 | 255 [バイト] |
| 1 インデクス型当たりの適用関数の数 | 30 [個] |
| サブタイプとして定義された抽象データ型の総数 | 0 [個] |
| 抽象データ型の総数 | SGMLTEXT は 1 個として加算 |

インデクスの格納ページ数の設定項目と設定値

| 設定項目 | 設定値 |
|----------------------------|--------------------------------------|
| 作成するユーザ定義データ型の数 | 1 [個] |
| ユーザ定義データ型名称の長さの平均値 | 8 [バイト] |
| 1 ユーザ定義データ型当たりの属性数の平均値 | 1 [個] |
| ユーザ定義データ型属性名称の長さの平均値 | 30 [バイト] |
| プラグイン名称の長さの平均値 | 8 [バイト] |
| 作成するプラグイン数 | 2 [個] |
| インデクス型名称の長さの平均値 | 8 [バイト] |
| プラグイン関数の数の平均値 | 30 [個] 複数ある場合は各プラグインの適用関数の数との平均 |
| プラグイン関数名称の長さの平均値 | 20 [バイト] |
| 1 プラグイン関数当たりのパラメタ数の平均値 | 6 [個] |
| 1 プラグイン関数当たりのパラメタ名称の長さの平均値 | 8 [バイト] |

(d) レジストリ用 RD エリア容量

Text Search Plug-in が使用する HiRDB のレジストリ用 RD エリアの容量は、次の値で見積もってください。レジストリ用 RD エリア容量の見積もり式は、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。

レジストリ管理表の格納ページ数

| 設定箇所 | 変数 | 設定値 |
|------------------|----|------------|
| レジストリ管理表のコンテキスト数 | a= | 3 |
| レジストリコンテキスト名長 | c= | 16 [バイト] |
| アクセスパスワード長 | d= | 12 [バイト] |

| 設定箇所 | 変数 | 設定値 |
|------------|----|----------------------------------|
| レジストリキー名称長 | f= | DTD 登録名称長 + 正規化パラメタ登録名称長 [バイト] |
| レジストリキー値長 | g= | DTD 長 + 正規化パラメタ長 [バイト] |

レジストリ管理表のインデクス格納ページ数

| 設定箇所 | 変数 | 設定値 |
|---------------|----|----------------------------------|
| レジストリコンテキスト名長 | a= | 16 [バイト] |
| レジストリキー名称長 | f= | DTD 登録名称長 + 正規化パラメタ登録名称長 [バイト] |

(e) レジストリ LOB 用 RD エリア容量

Text Search Plug-in が使用する HiRDB のレジストリ LOB 用 RD エリアの容量は、次の値で見積もってください。レジストリ LOB 用 RD エリア容量の見積もり式は、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。

| 設定箇所 | 変数 | 設定値 |
|-------------------------|-----|---------------------|
| 32,000 バイトを超えるレジストリキー値数 | b= | DTD と正規化パラメタの総数 |
| 32,000 バイトを超えるレジストリキー値長 | Cj= | DTD と正規化パラメタの長さの総合計 |

(2) FREEWORD プラグインおよび IXFREEWORD プラグインの場合

(a) ユーザ用 RD エリア容量

FREEWORD 型列を定義した表を格納するユーザ用 RD エリアの容量は、次の値で見積もってください。ユーザ用 RD エリアの見積もり式は、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。

| 設定箇所 | 変数 | 設定値 |
|---------|-----|---------------|
| 各列のデータ長 | di= | 次の説明を参照してください |

freewordsize rdareasize - 表データの合計 (各列のデータ長の総和) - 100 の場合
 $di [\text{バイト}] = \text{freewordsize} + 50$

freewordsize > rdareasize - 表データの合計 (各列のデータ長の総和) - 100 の場合
 $di [\text{バイト}] = 60$

この場合、表の格納ページ数 (ページ) に、次の値を加算してください。

$$\text{freewordsize} / \text{rdareasize} \times \text{freewordcnt}$$

2. Text Search Plug-in システムの構築

freewordsize : 登録文字列データ長 [バイト]
 freewordcnt : 登録文字列データ件数
 rdareasize : ユーザ用 RD エリアのページ長 [バイト]

(b) ユーザ LOB 用 RD エリア容量

HiRDB のユーザ LOB 用 RD エリアの容量の見積もりでは、次に説明する IXFREEWORD インデクス情報格納用 RD エリアの容量を計算します。

IXFREEWORD インデクス情報格納用 RD エリア

次の値を設定して、IXFREEWORD インデクス情報を格納するユーザ LOB 用 RD エリアの容量を見積もってください。分割遅延登録を使用する場合は、指定したインデクス分割遅延登録用領域サイズを加算したサイズとなります。また 1 つの RD エリアサイズの最大サイズは 64GB です。64GB を超える場合は表分割してください。ユーザ LOB 用 RD エリアの容量見積もり式は、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。

| 設定箇所 | 変数 | 設定値 |
|-----------------|-----|--|
| LOB 列の行の総数 | b= | 5 + 分割遅延登録を使用する場合 =1 分割遅延登録を使用しない場合 =0 |
| 各 BLOB データのデータ長 | Cj= | 次の説明を参照してください |

C1 のサイズ

$$C_1 = \max \left(\left(24 \times \text{freewordcnt} \right) / 8192 + \left(4 \times \text{freewordcnt} \right) / \left(512 \times 8192 \right) + \left(4 \times \text{freewordcnt} \right) / \left(1024 \times 8192 \right) \right) \times 8192 + 8192, 128000 + 8192)$$

C2 のサイズ

$$C_2 = 3750 \times 8192 + 8192$$

C3 のサイズ

通常インデクスの場合または英文検索用インデクス (ENGLISH) の場合

$$C_3 = \max \left(\left(\text{freewordcnt} \times \text{freewordsize} \times 3 \right) / 8192 \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

英文検索用インデクス (ENGLISH_STANDARD) を使用した場合

$$C_3 = \max \left(\left(\text{freewordcnt} \times \text{freewordsize} \times 5 \right) / 8192 \times 8192 + 8192, 200 \times 1024 \times 1024 \right)$$

C4 のサイズ

$$C_4 = 160 \times 8192 + 8192$$

C5 のサイズ

$$C_5 = 160 \times 8192 + 8192$$

C6 のサイズ

CREATEMODE=2 の分割遅延登録用インデックスの場合 (DELAY_SIZE 設定値のインデックス分割遅延登録用領域サイズ)

$$C_6 = C_1 \times 2 + 1024 \times 1024$$

ただし, C_1 中の freewordcnt は遅延登録文書数です。

CREATEMODE=3 の分割遅延登録・更新系 SQL 同時実行用インデックスの場合 (DELAY_SIZE 設定値の遅延登録時に情報を退避する領域サイズ)

$$C_6 = 16 \times \text{delayfreewordcnt} + 12 \times 8192$$

delayfreewordcnt : 遅延登録文書数

C7 のサイズ

UCS4 インデックスの場合 (初期作成)

$$C_7 = 50 \times 1024 \times 1024$$

UCS4 インデックスの場合

次のサイズを Available Work Area Size (使用可能ワーク領域サイズ) ¹ に確保してください。

$$C_7 = 200 \times 1024 \times 1024 + \text{freewordcnt} \times 1 \text{ 登録文字列中の UCS4 文字数 }^2 \times 5$$

注 1

phxidxls (インデックス情報取得ユーティリティ) を使って取得したインデックス情報に表示されます。HiRDB の pdmod コマンドによる RD エリアの拡張を行うと拡張されます。

注 2

UCS4 で表現できる文字の中で, UCS2 を超える範囲。

0x00010000 ~ 0x7fffffff

英語文字列検索用インデックスの場合

freewordcnt : 登録文字列データ件数

freewordsize : 平均登録文字列データ長 [バイト]

(c) データディクショナリの RD エリア容量

Text Search Plug-in で, FREEWORD 型と IXFREEWORD 型を表に適用する場合は, 次の値を設定して HiRDB のデータディクショナリ用 RD エリアの容量を見積もってください。データディクショナリ用 RD エリア容量の見積もり式は, 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。

2. Text Search Plug-in システムの構築

表の格納ページ数の設定項目と設定値

| 設定項目 | 設定値 |
|--------------------------|--------------------------|
| ユーザ定義型の名称長の平均値 | 8 [バイト] |
| プラグインオプションの平均値 | 80 [バイト] |
| 抽象データ型名称の長さの平均値 | 8 [バイト] |
| ユーザ定義型の総数 | 1 [個] |
| 1 データ型当たりの属性数の平均値 | 1 [個] |
| データ型識別子の長さの平均値 | 8 [バイト] |
| ユーザ定義型で定義された属性数 | 1 [個] |
| プラグインの総数 | 2 [個] |
| プラグイン名称の長さの平均値 | 8 [バイト] |
| 抽象データ型 / インデクス型名称の長さの平均値 | 8 [バイト] |
| インデクス型の総数 | 1 [個] |
| インデクス型識別子の長さの平均値 | 4 [バイト] |
| 抽象データ型名称の長さの平均値 | 8 [バイト] |
| プラグインインデクスの総数 | 1 [個] |
| インデクス型識別子の長さの平均値 | 4 [バイト] |
| 抽象データ型名称の長さの平均値 | 8 [バイト] |
| 抽象データ型で定義された属性の総数 | 1 [個] |
| プラグインインデクス適用関数名称の長さの平均値 | 10 [バイト] |
| 1 プラグインインデクス当たりの適用関数の数 | 30 [個] |
| 1 プラグイン当たりのパラメタ名称の長さの平均値 | 30 [バイト] |
| プラグインライブラリパス名称の長さの平均値 | 255 [バイト] |
| プラグインの注釈の長さの平均値 | 255 [バイト] |
| 1 インデクス型当たりの適用関数の数 | 30 [個] |
| サブタイプとして定義された抽象データ型の総数 | 0 [個] |
| 抽象データ型の総数 | FREEWORD は 1 [個] として加算 |

インデクスの格納ページ数の設定項目と設定値

| 設定項目 | 設定値 |
|------------------------|------------|
| 作成するユーザ定義データ型の数 | 1 [個] |
| ユーザ定義データ型名称の長さの平均値 | 8 [バイト] |
| 1 ユーザ定義データ型当たりの属性数の平均値 | 1 [個] |
| ユーザ定義データ型属性名称の長さの平均値 | 30 [バイト] |

| 設定項目 | 設定値 |
|---------------------------|-----------------------------------|
| プラグイン名称の長さの平均値 | 8 [バイト] |
| 作成するプラグイン数 | 2 [個] |
| インデクス型名称の長さの平均値 | 8 [バイト] |
| プラグイン関数の数の平均値 | 30 [個] 複数ある場合は各プラグインの適用関数の数との平均 |
| プラグイン関数名の長さの平均値 | 20 [バイト] |
| 1 プラグイン関数当たりのパラメタ数の平均値 | 6 [個] |
| 1 プラグイン関数当たりのパラメタ名の長さの平均値 | 8 [バイト] |

(d) レジストリ用 RD エリア容量

Text Search Plug-in で FREEWORD 型と IXFREEWORD インデクス型を使用する場合 HiRDB のレジストリ用 RD エリアは使用しません。

(e) レジストリ LOB 用 RD エリア容量

Text Search Plug-in で FREEWORD 型と IXFREEWORD インデクス型を使用する場合 HiRDB のレジストリ用 RD エリアは使用しません。

2.3.3 ファイル容量

Text Search Plug-in のファイル容量の見積もり方法について説明します。見積もりの必要なファイル容量は、次に説明する更新作業用ファイルと、同義語辞書ファイルおよびシステムログファイルの容量です。

(1) SGML プラグインおよび n-gram インデクスプラグインの場合

(a) 同義語辞書ファイル

Text Search Plug-in は、同義語辞書ファイルを次の場所に格納します。

UNIX の場合

```
$PDDIR/plugin/_phsgml/dic
$PDDIR/plugin/_phsgml/dic_c
$PDDIR/plugin/_phsgml/dic_w
$PDDIR/plugin/_phsgml/dic_y
$PDDIR/plugin/_phsgml/dic_cw
$PDDIR/plugin/_phsgml/dic_cy
$PDDIR/plugin/_phsgml/dic_wy
$PDDIR/plugin/_phsgml/dic_cwy
$PDDIR/plugin/_phsgml/dic_cd
$PDDIR/plugin/_phsgml/dic_cdr
$PDDIR/plugin/_phsgml/dic_cr
$PDDIR/plugin/_phsgml/dic_cwd
$PDDIR/plugin/_phsgml/dic_cwdr
$PDDIR/plugin/_phsgml/dic_cwr
```

2. Text Search Plug-in システムの構築

```
$PDDIR/plugin/_phsgml/dic_cwyd
$PDDIR/plugin/_phsgml/dic_cwydr
$PDDIR/plugin/_phsgml/dic_cwyr
$PDDIR/plugin/_phsgml/dic_cyd
$PDDIR/plugin/_phsgml/dic_cydr
$PDDIR/plugin/_phsgml/dic_cyr
$PDDIR/plugin/_phsgml/dic_d
$PDDIR/plugin/_phsgml/dic_dr
$PDDIR/plugin/_phsgml/dic_r
$PDDIR/plugin/_phsgml/dic_wd
$PDDIR/plugin/_phsgml/dic_wdr
$PDDIR/plugin/_phsgml/dic_wr
$PDDIR/plugin/_phsgml/dic_wyd
$PDDIR/plugin/_phsgml/dic_wydr
$PDDIR/plugin/_phsgml/dic_wyr
$PDDIR/plugin/_phsgml/dic_yd
$PDDIR/plugin/_phsgml/dic_ydr
$PDDIR/plugin/_phsgml/dic_yr
```

Windows の場合

```
%PDDIR%\plugin\_phsgml\dic
%PDDIR%\plugin\_phsgml\dic_c
%PDDIR%\plugin\_phsgml\dic_w
%PDDIR%\plugin\_phsgml\dic_y
%PDDIR%\plugin\_phsgml\dic_cw
%PDDIR%\plugin\_phsgml\dic_cy
%PDDIR%\plugin\_phsgml\dic_wy
%PDDIR%\plugin\_phsgml\dic_cwy
%PDDIR%\plugin\_phsgml\dic_cd
%PDDIR%\plugin\_phsgml\dic_cdr
%PDDIR%\plugin\_phsgml\dic_cr
%PDDIR%\plugin\_phsgml\dic_cwd
%PDDIR%\plugin\_phsgml\dic_cwdr
%PDDIR%\plugin\_phsgml\dic_cwr
%PDDIR%\plugin\_phsgml\dic_cwyd
%PDDIR%\plugin\_phsgml\dic_cwydr
%PDDIR%\plugin\_phsgml\dic_cwyr
%PDDIR%\plugin\_phsgml\dic_cyd
%PDDIR%\plugin\_phsgml\dic_cydr
%PDDIR%\plugin\_phsgml\dic_cyr
%PDDIR%\plugin\_phsgml\dic_d
%PDDIR%\plugin\_phsgml\dic_dr
%PDDIR%\plugin\_phsgml\dic_r
%PDDIR%\plugin\_phsgml\dic_wd
%PDDIR%\plugin\_phsgml\dic_wdr
%PDDIR%\plugin\_phsgml\dic_wr
%PDDIR%\plugin\_phsgml\dic_wyd
%PDDIR%\plugin\_phsgml\dic_wydr
%PDDIR%\plugin\_phsgml\dic_wyr
%PDDIR%\plugin\_phsgml\dic_yd
%PDDIR%\plugin\_phsgml\dic_ydr
%PDDIR%\plugin\_phsgml\dic_yr
```

容量は、次に示す式で計算します。

同義語辞書容量 [バイト] = (wordnum × (20 + wordlen) + 1000) × dicnum

- wordnum : 登録語数
- wordlen : 平均語長 [バイト]
- dicnum : 登録辞書数

(b) 概念検索用統計情報サイズ (1RD エリア当たりの phnconget 結果)

収集モード指定 : normal

概念検索用統計情報サイズ [バイト] = phnidxls で表示された概念統計情報サイズ
(Size of Original Concept File) × 8192

収集モード指定 : all

概念検索用統計情報サイズ [バイト] = phnidxls で表示された概念統計情報サイズ
(Size of Original Concept File) × 8192 × 1.5

(c) 概念検索用統合統計情報サイズ (phnconbuild 結果)

概念検索用統合統計情報サイズ [バイト] = (各概念検索用統計情報サイズ) + MAX
(各概念検索用統計情報サイズ)

(d) ログ量

ログ量は、既登録文書量や登録する文書を構成する文字列に依存します。ここでは平均値で記載します。

SGMLTEXT 型列の追加

jnl [キロバイト] = 32

PURGE TABLE

jnl [キロバイト] = 48

これ以降の INSERT 時、DELETE 時、UPDATE 時、一括更新時およびインクリメンタル処理のログ取得には、ログ取得モード、更新前ログ取得モードおよびログレスモードが指定できます。ただし、ログ取得モードでは、更新前ログ取得モードの約 2 倍必要になります。ここでは更新前ログ取得モードで記載します。

INSERT 時

jnl [キロバイト]
= (docsize × 3 × 10⁻³) / 4 + 1 × 4 + (60 × docsize × 10⁻³) / 4 × 4

docsize : 平均登録文書サイズ [バイト]

DELETE 時

jnl [キロバイト] = 40

UPDATE 時

jnl [キロバイト] = INSERT で必要分 + DELETE で必要分

2. Text Search Plug-in システムの構築

一括更新時 (pload, pdrorg, phnmerge, CREATE INDEX, startidx)

$$\begin{aligned} \text{jnl [キロバイト]} \\ = & (\text{docsize} \times 3 \times 10^{-3}) / 4 + 1 \quad \times 4 \times \text{doccnt} \\ & + (30 \times \text{docsize} \times 10^{-3} \times \text{doccnt}) / 4 \quad \times 4 \end{aligned}$$

docsize : 平均登録文書サイズ [バイト]

doccnt : 登録文書数

インクリメンタル処理

$$\begin{aligned} \text{jnl [キロバイト]} \\ = & (\text{余剰インデクス容量} \times 8 \times 103) / \text{インクリメント対象トライノード数} \\ & \times \text{インクリメント対象トライノード数} \times 10 \end{aligned}$$

インクリメンタル処理で使用する余剰インデクス容量およびインクリメント対象トライノード数については、「phnidxls (インデクス情報の取得ユティリティ)」で取得できます。

インデクス詰め替え処理

$$\text{jnl [キロバイト]} = \text{インデクスデータ量}$$

統計情報反映処理

$$\text{jnl [キロバイト]} = \text{反映する統計情報サイズ} + 24$$

(2) FREWORD プラグインおよび IXFREWORD プラグインの場合

(a) 同義語辞書ファイル

Text Search Plug-in は、同義語辞書ファイルを次の場所に格納します。

UNIX の場合

```
$PDDIR/plugin/_phcfwd/dic
$PDDIR/plugin/_phcfwd/dic_c
$PDDIR/plugin/_phcfwd/dic_w
$PDDIR/plugin/_phcfwd/dic_y
$PDDIR/plugin/_phcfwd/dic_cw
$PDDIR/plugin/_phcfwd/dic_cy
$PDDIR/plugin/_phcfwd/dic_wy
$PDDIR/plugin/_phcfwd/dic_cwy
$PDDIR/plugin/_phcfwd/dic_cd
$PDDIR/plugin/_phcfwd/dic_cdr
$PDDIR/plugin/_phcfwd/dic_cr
$PDDIR/plugin/_phcfwd/dic_cwd
$PDDIR/plugin/_phcfwd/dic_cwdr
$PDDIR/plugin/_phcfwd/dic_cwr
$PDDIR/plugin/_phcfwd/dic_cwyd
$PDDIR/plugin/_phcfwd/dic_cwydr
$PDDIR/plugin/_phcfwd/dic_cwyr
$PDDIR/plugin/_phcfwd/dic_cyd
$PDDIR/plugin/_phcfwd/dic_cydr
```

```

$PDDIR/plugin/_phcfwd/dic_cyr
$PDDIR/plugin/_phcfwd/dic_d
$PDDIR/plugin/_phcfwd/dic_dr
$PDDIR/plugin/_phcfwd/dic_r
$PDDIR/plugin/_phcfwd/dic_wd
$PDDIR/plugin/_phcfwd/dic_wdr
$PDDIR/plugin/_phcfwd/dic_wr
$PDDIR/plugin/_phcfwd/dic_wyd
$PDDIR/plugin/_phcfwd/dic_wydr
$PDDIR/plugin/_phcfwd/dic_wyr
$PDDIR/plugin/_phcfwd/dic_yd
$PDDIR/plugin/_phcfwd/ic_ydr
$PDDIR/plugin/_phcfwd/dic_yr

```

Windows の場合

```

%PDDIR%\plugin\_%phcfwd%\dic
%PDDIR%\plugin\_%phcfwd%\dic_c
%PDDIR%\plugin\_%phcfwd%\dic_w
%PDDIR%\plugin\_%phcfwd%\dic_y
%PDDIR%\plugin\_%phcfwd%\dic_cw
%PDDIR%\plugin\_%phcfwd%\dic_cy
%PDDIR%\plugin\_%phcfwd%\dic_wy
%PDDIR%\plugin\_%phcfwd%\dic_cwy
%PDDIR%\plugin\_%phcfwd%\dic_cd
%PDDIR%\plugin\_%phcfwd%\dic_cdr
%PDDIR%\plugin\_%phcfwd%\dic_cr
%PDDIR%\plugin\_%phcfwd%\dic_cwd
%PDDIR%\plugin\_%phcfwd%\dic_cwdr
%PDDIR%\plugin\_%phcfwd%\dic_cwr
%PDDIR%\plugin\_%phcfwd%\dic_cwyd
%PDDIR%\plugin\_%phcfwd%\dic_cwydr
%PDDIR%\plugin\_%phcfwd%\dic_cwyr
%PDDIR%\plugin\_%phcfwd%\dic_cyd
%PDDIR%\plugin\_%phcfwd%\dic_cydr
%PDDIR%\plugin\_%phcfwd%\dic_cyr
%PDDIR%\plugin\_%phcfwd%\dic_d
%PDDIR%\plugin\_%phcfwd%\dic_dr
%PDDIR%\plugin\_%phcfwd%\dic_r
%PDDIR%\plugin\_%phcfwd%\dic_wd
%PDDIR%\plugin\_%phcfwd%\dic_wdr
%PDDIR%\plugin\_%phcfwd%\dic_wr
%PDDIR%\plugin\_%phcfwd%\dic_wyd
%PDDIR%\plugin\_%phcfwd%\dic_wydr
%PDDIR%\plugin\_%phcfwd%\dic_wyr
%PDDIR%\plugin\_%phcfwd%\dic_yd
%PDDIR%\plugin\_%phcfwd%\dic_ydr
%PDDIR%\plugin\_%phcfwd%\dic_yr

```

容量は、次に示す式で計算します。

同義語辞書容量 [バイト] = (wordnum × (20 + wordlen) + 1000) × dicnum

wordnum : 登録語数

wordlen : 平均語長 [バイト]

2. Text Search Plug-in システムの構築

dicnum : 登録辞書数

(b) ログ量

ログ量は、既登録文字列データ量や登録する文字列の内容に依存します。ここでは平均値で記載します。

FREEWORD 型列の追加

$$\text{jnl [キロバイト]} = 32$$

PURGE TABLE

$$\text{jnl [キロバイト]} = 48$$

これ以降の INSERT 時, DELETE 時, UPDATE 時, 一括更新時およびインクリメンタル処理のログ取得には、ログ取得モード, 更新前ログ取得モードおよびログレスモードが指定できます。ただし、ログ取得モードでは、更新前ログ取得モードの約 2 倍必要になります。ここでは更新前ログ取得モードで記載します。

INSERT 時

$$\begin{aligned} \text{jnl [キロバイト]} \\ = & (\text{freewordsize} \times 3 \times 10^{-3}) / 4 + 1 \times 4 \\ & + (60 \times \text{freewordsize} \times 10^{-3}) / 4 \times 4 \end{aligned}$$

freewordsize : 平均登録文字列データ長 [バイト]

DELETE 時

$$\text{jnl [キロバイト]} = 40$$

UPDATE 時

$$\text{jnl [キロバイト]} = \text{INSERT で必要分} + \text{DELETE で必要分}$$

一括更新時 (pdload, pdrorg, phnmerge, CREATE INDEX)

$$\begin{aligned} \text{jnl [キロバイト]} \\ = & (\text{freewordsize} \times 3 \times 10^{-3}) / 4 + 1 \times 4 \times \text{freewordcnt} \\ & + (30 \times \text{freewordsize} \times 10^{-3} \times \text{doccnt}) / 4 \times 4 \end{aligned}$$

freewordsize : 平均登録文字列データ長 [バイト]

freewordcnt : 登録文字列データ件数

doccnt : 登録文書数

インクリメンタル処理

$$\begin{aligned} \text{jnl [キロバイト]} \\ = & (\text{余剰インデクス容量} \times 8 \times 103) / \text{インクリメント対象トライノード数} \\ & \times \text{インクリメント対象トライノード数} \times 10 \end{aligned}$$

インクリメンタル処理で使用する余剰インデクス容量およびインクリメント対象トライノード数については、「phxidxls（インデクス情報の取得ユティリティ）」で取得できます。

インデクス詰め替え処理

jnl [キロバイト] = インデクスデータ量

2.3.4 論理ファイル数

次に示す Text Search Plug-in が使用する論理ファイル数の値は、HiRDB の排他資源数の見積もりで使われます。排他資源数の見積もり方法についてはマニュアル「HiRDB システム定義」を参照してください。

| プラグイン | 論理ファイル数 | ファイル作成場所 |
|-------------------|--|------------------------------------|
| SGML プラグイン | 2 [個] | SGMLTEXT データ格納用のユーザ LOB 用 RD エリア |
| n-gram インデクスプラグイン | 5 + + + + + [個] RD エリア容量が 2 ギガバイト以下の場合 =0 RD エリア容量が 2 ギガバイトを超える場合 =1 ~ 15 差分インデクスを使用しない場合 =0 差分インデクスを使用する場合 =4 分割遅延登録を使用する場合 =5 分割遅延登録を使用しない場合 =0 統計情報を反映する場合 =1 統計情報を反映しない場合 =0 インデクスファイル自動拡張を使用する場合 =62 インデクスファイル自動拡張を使用しない場合 =0 | n-gram インデクス情報格納用のユーザ LOB 用 RD エリア |
| FREWORD プラグイン | 0 [個] | 作成しません |

2. Text Search Plug-in システムの構築

| プラグイン | 論理ファイル数 | ファイル作成場所 |
|------------------|--|--|
| IXFREEWORD プラグイン | 5 + + + [個] RD エリア容量が 2 ギガバイト以下の場合 =0 RD エリア容量が 2 ギガバイトを超える場合 =1 ~ 15 分割遅延登録を使用する場合 =5 分割遅延登録を使用しない場合 =0 インデクスファイル自動拡張を使用する場合 =62 インデクスファイル自動拡張を使用しない場合 =0 | IXFREEWORD インデクス情報格納用のユーザ LOB 用 RD エリア |

2.4 Text Search Plug-in の環境設定

この節では、Text Search Plug-in のセットアップ・登録、レジストリ機能の初期設定、およびシステム共通定義の変更など、Text Search Plug-in の環境設定について説明します。セットアップは UNIX の場合にだけ必要です。Windows の場合は、インストール時にセットアップが実行されますので、「2.4.2 Text Search Plug-in の登録」から参照してください。

なお、Text Search Plug-in の環境設定では、HiRDB が提供しているユティリティおよび運用コマンドを使用する場合があります。これらのユティリティおよび運用コマンドの使用方法については、マニュアル「HiRDB コマンドリファレンス」および「HiRDB システム導入・設計ガイド」を参照してください。なお、ご使用の OS が AIX の場合は、マニュアル「HiRDB システム導入・設計ガイド」を参照して、AIX に固有の指定を設定してください。

2.4.1 Text Search Plug-in のセットアップ (UNIX の場合)

セットアップとは、プラグイン実装ファイルを HiRDB 運用ディレクトリにコピーすることを指します。文書の全文検索機能や概念検索機能を使う場合、文書検索プラグイン (SGML プラグインおよび n-gram インデックスプラグイン) に対してセットアップが必要です。また、文字列データの全文検索機能を使う場合、文字列検索プラグイン (FREEWORD プラグインおよび IXFREEWORD プラグイン) に対して、セットアップが必要です。

Text Search Plug-in のセットアップには、HiRDB の運用コマンド「pdplgset (プラグインのセットアップ)」を使用します。次の形式で実行してください。

コマンドの形式

```
pdplgset プラグイン名 プラグインインストールディレクトリ
```

引数

プラグイン名

セットアップするプラグインの名称を指定します。Text Search Plug-in では、各プラグインに対応してプラグイン名を次のように指定してください。

| 文書検索プラグイン | プラグイン名 |
|--------------------|----------|
| SGML プラグイン | _phsgml |
| n-gram インデックスプラグイン | _phngram |

| 文字列検索プラグイン | プラグイン名 |
|------------------|----------|
| FREEWORD プラグイン | _phcfwd |
| IXFREEWORD プラグイン | _phxfwrđ |

プラグインインストールディレクトリ

サーバマシンにインストールされた、プラグインがあるディレクトリを指定します。Text Search Plug-in では、OS に対応してプラグインインストールディレクトリを /opt/TSPlugin に指定してください。

コマンド実行例

```
pdplgset _phsgml /opt/TSPlugin
pdplgset _phngram /opt/TSPlugin
pdplgset _phcfwd /opt/TSPlugin
pdplgset _phxfwrđ /opt/TSPlugin
```

注意事項

- すでに、同一ディレクトリにプラグインがセットアップされている場合、HiRDB が稼働中でないかを確認してください。稼働している場合、HiRDB の運用コマンド「pdstop (HiRDB システム・ユニット・サーバの終了)」で停止させたあと、「pdplgset (プラグインのセットアップ)」で -d オプションを指定し、プラグインを HiRDB からアンセットアップしてください。
- プラグインのアンセットアップ (pdplgset -d) を行う前にプラグインの固有ファイルを退避させる必要があります。詳細は「5.3.2(2) 運用環境のバックアップ」を参照してください。

2.4.2 Text Search Plug-in の登録

Text Search Plug-in の登録とは、プラグインに関する定義情報をディクショナリへ登録すること、抽象データ型定義またはインデクス型定義を実行することを指します。Text Search Plug-in で提供する文書の全文検索機能や概念検索機能を使う場合、文書検索プラグイン (SGML プラグインおよび n-gram インデクスプラグイン) に対して登録が必要です。また、文字列データの全文検索機能を使う場合、文字列検索プラグイン (FREEWORD プラグインおよび IXFREEWORD プラグイン) に対して登録が必要です。

なお、文書検索プラグインを使用する場合、SGML プラグインおよび n-gram インデクスプラグインの両方に対して、登録を実行してください。文字列検索プラグインを使用する場合、FREEWORD プラグインおよび IXFREEWORD プラグインの両方に対して、登録を実行してください。

また、HiRDB がパラレルサーバの場合、登録およびバージョンアップは任意のサーバで実行してください。

(1) Text Search Plug-in のバージョンアップ

Text Search Plug-in のバージョンアップには、HiRDB の運用コマンド「pdplgrgst (プラグインの登録・削除)」を使用します。バージョンアップの手順については、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。

UNIX 使用時の注意

なお、バージョンアップ時に新たにプラグインを追加して使用する場合は、追加するプラグインについてセットアップから実行してください。

(2) Text Search Plug-in の登録

Text Search Plug-in の登録には、HiRDB の運用コマンド「pdplgrgst (プラグインの登録・削除)」を使用します。登録の手順については、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。

Text Search Plug-in の登録の場合は、次の形式で実行してください。なお、文書検索プラグインを使用する場合は、SGML プラグイン、n-gram インデクスプラグインの順に登録してください。また、文字列検索プラグインを使用する場合は、FREEWORD プラグイン、IXFREEWORD プラグインの順に登録してください。

コマンドの形式

```
pdplgrgst プラグイン定義ファイル名 PICファイル名
```

引数

プラグイン定義ファイル名

プラグイン定義ファイルのファイル名を指定します。コマンドを入力するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定します。Text Search Plug-in では、各プラグインに対応してプラグイン定義ファイル名を次のように指定してください。

| 文書検索プラグイン | プラグイン定義ファイル名 |
|-------------------|--------------|
| SGML プラグイン | _phsgml.adt |
| n-gram インデクスプラグイン | _phngram.idx |

| 文字列検索プラグイン | プラグイン定義ファイル名 |
|------------------|--------------|
| FREEWORD プラグイン | _phcfwd.adt |
| IXFREEWORD プラグイン | _phxfwr.idx |

PIC ファイル名

PIC ファイルのファイル名を指定します。コマンドを入力するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定しま

2. Text Search Plug-in システムの構築

す。Text Search Plug-in では、各プラグインに対応して PIC ファイル名を次のように指定してください。

| 文書検索プラグイン | PIC ファイル名 |
|-------------------|--------------|
| SGML プラグイン | _phsgml.pic |
| n-gram インデクスプラグイン | _phngram.pic |

| 文字列検索プラグイン | PIC ファイル名 |
|------------------|--------------|
| FREEWORD プラグイン | _phcfwd.pic |
| IXFREEWORD プラグイン | _phxfwrд.pic |

コマンド実行例 (UNIX の場合)

```
pdplgrgst $PDDIR/plugin/_phsgml/etc/_phsgml.adt ¥
           $PDDIR/plugin/_phsgml/etc/_phsgml.pic
pdplgrgst $PDDIR/plugin/_phngram/etc/_phngram.idx ¥
           $PDDIR/plugin/_phngram/etc/_phngram.pic
pdplgrgst $PDDIR/plugin/_phcfwd/etc/_phcfwd.adt ¥
           $PDDIR/plugin/_phcfwd/etc/_phcfwd.pic
pdplgrgst $PDDIR/plugin/_phxfwrд/etc/_phxfwrд.idx ¥
           $PDDIR/plugin/_phxfwrд/etc/_phxfwrд.pic
```

上記は、2 行でコマンドを入力した例を記載していますが、1 行で入力する場合は、「¥」(エスケープ文字)を入力しないでください。

コマンド実行例 (Windows の場合)

```
pdplgrgst %PDDIR%¥plugin¥_phsgml¥etc¥_phsgml.adt
           %PDDIR%¥plugin¥_phsgml¥etc¥_phsgml.pic
pdplgrgst %PDDIR%¥plugin¥_phngram¥etc¥_phngram.idx
           %PDDIR%¥plugin¥_phngram¥etc¥_phngram.pic
pdplgrgst %PDDIR%¥plugin¥_phcfwd¥etc¥_phcfwd.adt
           %PDDIR%¥plugin¥_phcfwd¥etc¥_phcfwd.pic
pdplgrgst %PDDIR%¥plugin¥_phxfwrд¥etc¥_phxfwrд.idx
           %PDDIR%¥plugin¥_phxfwrд¥etc¥_phxfwrд.pic
```

- 「PDDIR」は、HiRDB インストール先フォルダパスが格納された環境変数を示します。環境変数 PDDIR は「%」で囲んで「%PDDIR%」と入力してください。
- 上記の例では、コマンドの引数を改行して記述していますが、実際は改行しないで入力してください。

2.4.3 レジストリ機能の初期設定

Text Search Plug-in の文書検索プラグインでは、HiRDB のレジストリ機能を使用するため、レジストリ機能の初期設定が必要です。レジストリ機能の初期設定には、HiRDB

の「pdreginit (レジストリ機能初期設定ユーティリティ)」を使用して、レジストリ用 RD エリアおよびレジストリ LOB 用 RD エリアを作成します。

ただし、すでにほかのプラグインでレジストリ機能の初期設定をしている場合は、この操作は不要です。また、文字列検索プラグインでは、HiRDB のレジストリ機能を使用しないため、レジストリの初期設定をする必要はありません。

2.4.4 システム共通定義の変更

登録した Text Search Plug-in を使用できる状態にするため、一度、HiRDB の運用コマンド「pdstop (HiRDB システム・ユニット・サーバの終了)」で HiRDB を正常終了させたあと、すべてのサーバマシン上のシステム共通定義「pdsys」に、使用するプラグインを宣言するオペランド「pdplugin」を追加します。

システム共通定義「pdsys」の詳細については、マニュアル「HiRDB システム定義」を参照してください。

システム共通定義のオペランド「pdplugin」は、次の形式で指定してください。

オペランドの指定形式

```
pdplugin -n プラグイン名
```

引数

-n プラグイン名

使用するプラグインの名称を指定します。Text Search Plug-in では、各プラグインに対応してプラグイン名を次のように指定してください。

| 文書検索プラグイン | プラグイン名 |
|-------------------|----------|
| SGML プラグイン | _phsgml |
| n-gram インデクスプラグイン | _phngram |

| 文字列検索プラグイン | プラグイン名 |
|------------------|---------|
| FREEWORD プラグイン | _phcfwd |
| IXFREEWORD プラグイン | _phxfwr |

オペランド指定例

```
pdplugin -n _phsgml
pdplugin -n _phngram
pdplugin -n _phcfwd
pdplugin -n _phxfwr
```

プラグインを削除する場合には、HiRDB のシステム共通定義「pdsys」に追加した、使用するプラグインを宣言する「pdplugin」オペランドを削除します。このとき、このオ

2. Text Search Plug-in システムの構築

ペラントを追加しているすべてのサーバマシン上のシステム共通定義から削除してください。

2.5 文書検索プラグインのデータベースの作成

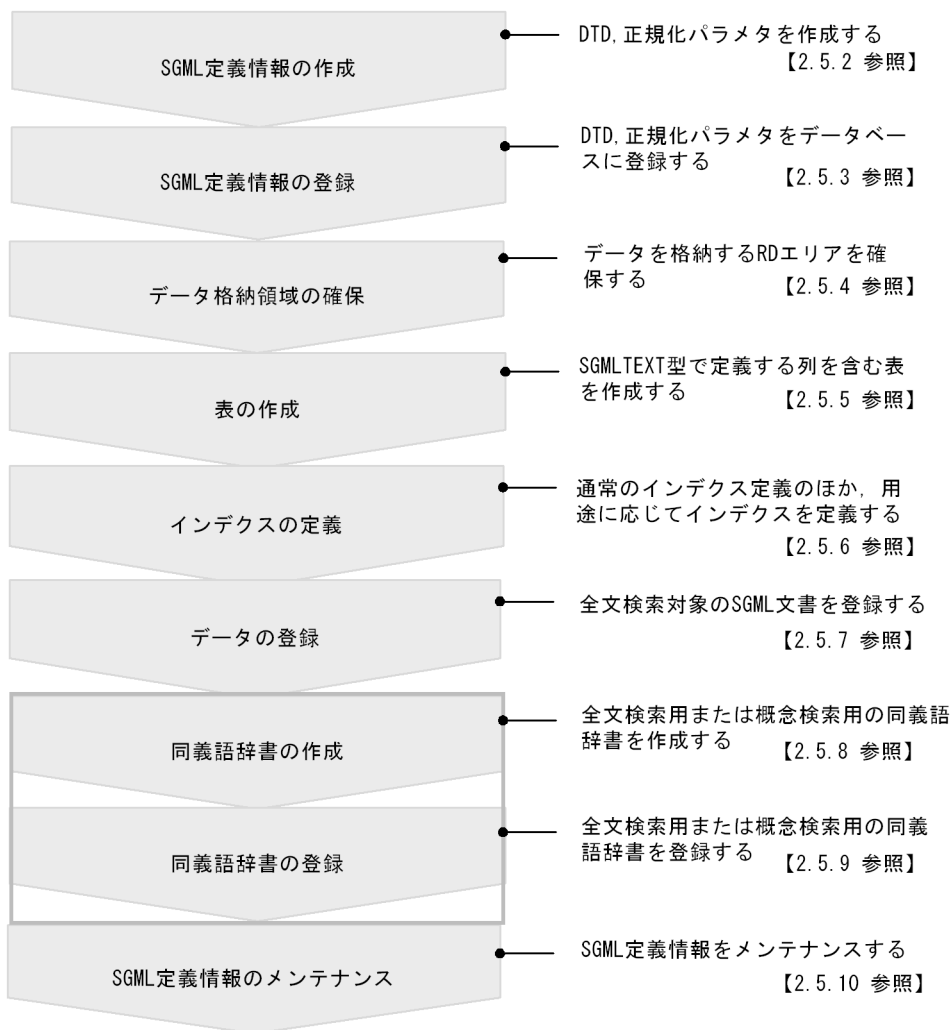
この節では、文書検索プラグインのデータベースの作成について説明します。

なお、Text Search Plug-in で必要となるデータベースを作成するには、HiRDB が提供しているユティリティおよび運用コマンドを使用する場合があります。これらのユティリティおよび運用コマンドの使用方法については、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。また、Text Search Plug-in のユティリティの使用方法については、「3. Text Search Plug-in のユティリティ」を参照してください。

2.5.1 文書検索プラグインのデータベース作成の流れ

Text Search Plug-in の導入後、表作成からデータベースへの登録までの操作方法の流れを次の図に示します。

図 2-2 データベース作成の流れ (文書検索プラグイン)



□ の内容は、表の定義から同義語展開検索機能を実行するまでの間に実行してください。

以降、上記の流れに従ってデータベースの作成方法について説明します。

2.5.2 SGML 定義情報の作成

Text Search Plug-in システムで文書検索プラグインの環境を構築するには、文書検索プラグインが使用する表の SGMLTEXT 型の列の作成に必要な情報を定義するファイルを作成します。必要な情報とは、SGML テキストデータの構造と文書要素を表すタグ名称などを定義する DTD、検索に不要なタグや文書要素の削除、参照タグと外字コードの対応を定義する正規化パラメタのことです。これらの情報をまとめて、SGML 定義情報と

います。

ファイルは、テキストエディタなどを使用して作成します。作成する列の種別によって必要なファイルは異なります。作成する列の種別ごとに必要となるファイルを次の表に示します。

表 2-4 文書検索プラグインの SGML 定義情報の作成に必要なファイル

| 必要なファイル名 | 列種別 | |
|-------------|-----------|---------|
| | 構造なし文書用の列 | 構造文書用の列 |
| DTD ファイル | × | |
| 正規化パラメタファイル | × | |

(凡例)

- : 必要
- × : 不要
- : 任意

注意事項

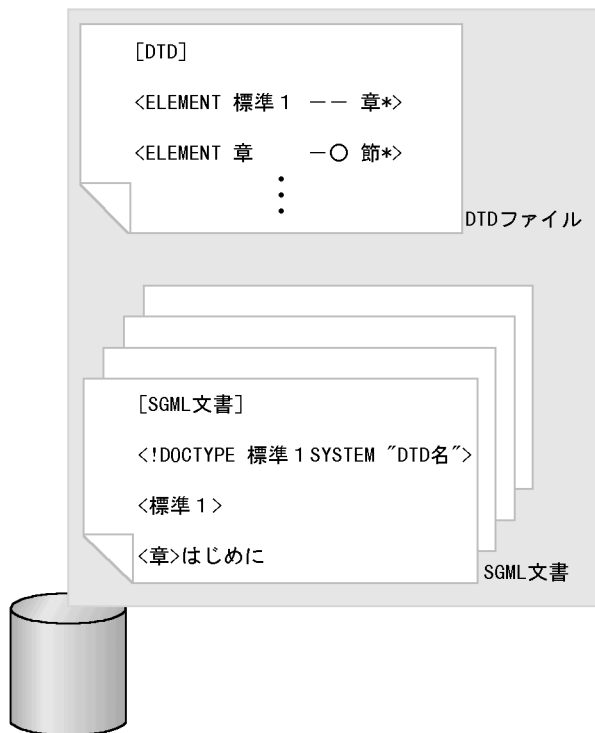
- Text Search Plug-in の定義に必要なファイルの名称は、ご利用の OS の制限に従ってください。
- 各ファイルの内容は、HiRDB をセットアップした文字コード、および「付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字」に示す文字コードで記述してください。

(1) DTD ファイル

DTD ファイルとは、構造文書用の列に対して、登録する文書の構造を定義するためのファイルです。一つの列に一つの DTD ファイルを作成します。DTD ファイルはテキストエディタなどで作成します。なお、構造なし文書用の列を作成する場合、このファイルを作成する必要はありません。

Text Search Plug-in では、一つの列に対して一つの DTD を対応付けます。通常、SGML 文書には DTD が含まれています。しかし、Text Search Plug-in で管理する列には、DTD は文書と切り離して入力します。これによって、DTD の異なる文書の登録を避けられます。列での DTD と SGML 文書の分割について、次の図に示します。

図 2-3 DTD と SGML 文書の分割



なお、列に DTD ファイルが登録されていれば、SGML 文書に DTD が含まれていても問題はありません。ただし、SGML 文書に含まれている DTD の内容が、列に登録されている DTD の内容と矛盾している場合は、エラーになります。

(a) Text Search Plug-in での SGML 文書の制限

Text Search Plug-in では SGML に関して次の制限があります。DTD を決定する場合は、これらの制限に注意してください。

- SGML 宣言指定は不要です。SGML 宣言が指定されていても、内容は無効です。
- DOCTYPE は最上位構造名になります。
- 一つの列には同じ DOCTYPE を持つ文書を登録してください。
- 公開識別子 (PUBLIC) は指定できません。
- システム識別子 (SYSTEM) には DTD 名称を指定してください。
- 外部エンティティは扱えません。
- 登録する SGML 文書は、次に示す文字コードで作成してください。
 - HiRDB をセットアップした文字コード
 - 「付録 A 登録する文書 (テキストデータ) および文字列データに使用できる文字」に示す文字コード
- SGML 文書中の一つのタグに対するネストレベルは 100 までです。

(b) SGML 文書中の改行コードの変換方法

SGML 文書中の改行コードの変換方法を次の表に示します。

表 2-5 SGML 文書中の改行コードの変換方法

| 改行コードの前の文字 | 改行コードの後ろの文字 | 変換方法 |
|--------------|--------------|---------------|
| 1 バイト文字 | 1 バイト文字 | 改行コードをスペースに置換 |
| | 1 バイト文字以外の文字 | |
| 1 バイト文字以外の文字 | 1 バイト文字 | 改行コードを削除 |
| | 1 バイト文字以外の文字 | |

そのほかの改行コードの変換規則は次のとおりです。

- 文書 (CDATA) の初めおよび終わりの改行コードは削除します。
- 改行コードだけの文字列は、一つの半角スペースに置き換えます。
- 改行コードが連続する場合は、一つの改行コードとして扱います。

(2) 正規化パラメタファイル

正規化パラメタファイルには、SGML 文書中のタグを制御するパラメタを記述します。正規化パラメタファイルはテキストエディタなどで作成します。なお、SGML 文書中のタグの制御が必要ない場合、このファイルを作成する必要はありません。

正規化パラメタファイルは、DTD 単位に作成します。つまり、構造文書用の列に対して一つ作成します。

(a) 正規化パラメタファイルの構成内容

正規化パラメタファイルは、正規化パラメタと特定文字データ変換マップで構成されます。

正規化パラメタ

- 接続対象要素リスト
要素型名のリストです。非構造要素のうち、前後にある文字データを接続する要素を指定します。ソース上では、指定した開始タグと対応する終了タグを削除する指定です。
- 削除対象要素リスト
要素型名のリストです。非構造要素のうち、内部にある下位要素や文字データを含めて、削除する要素を指定します。ソース上では、指定したタグ、構造内に含まれるタグおよび内容 (文字データ) をすべて削除する指定です。

特定文字データ変換マップ

ある特定の文字データを、共通的な文字コードへ変換するための指定です。SGML の構文上、テキスト中に直接記述できない文字 (例えば "<" など) や外字などの変換方法もこのマップに含めます。

(b) 正規化パラメタファイルの記述形式

正規化パラメタの指定方法

正規化パラメタは、要素型名に対して実行するオペランドを指定します。

記述形式

```
[ ELEMENT LIST ] <改行>  
"要素型名" {CONN|DELT}<改行>
```

オペランド

CONN：接続対象要素リスト

DELT：削除対象要素リスト

注意事項

- テキスト形式で作成してください。
- 要素型名は重複して指定できません。
- <改行> は改行コードを表しています。

特定文字データ変換マップの指定方法

特定文字データ変換マップは次の形式で記述します。

記述形式

```
[ CHANGE CHARACTER LIST ] <改行>  
共通的な文字コードへ変換する SGML 文書中の文字データ 変換する文字 <改行>
```

注意事項

- 共通的な文字コードへ変換する特定の文字データは、DTD ファイルに特定文字の実体を示すキーワードとして、"SDATA" で宣言されている必要があります。
- SDATA は文字コードを定義していませんので、全文検索を実行するために文字コードを定義する必要があります。
- テキスト形式で作成してください。
- SGML 文書中の値と変換する文字は 1 行で記述してください。
- 変換する文字に、改行コード (0x0a) は指定できません。
- <改行> は改行コードを表しています。

(c) 正規化パラメタファイルの記述例

正規化パラメタファイルの記述例を次の図に示します。

図 2-4 正規化パラメタファイルの記述例

```
[ELEMENT△LIST] <改行>
"hp" CONN<改行>
"secret" DELT<改行>

[CHANGE△CHARACTER△LIST] <改行>
{star} △☆<改行>
```

(凡例)

<改行> : 改行コード

(3) SGML 文書の正規化の例

SGML 文書の正規化の例を次の図に示します。

図 2-5 SGML 文書の正規化の例

● SGML文書

```
<!DOCTYPE body SYSTEM "DTD名">
<body>
<p>
近年、急速に広がりを見せるインターネットに
関する基礎講座を <hp>&star:〇月×日夕方6時
</hp>より、<hp>本社第一会議室</hp>にて開催
いたします。
<note>
なお、閉会後は懇親会を予定しております。差
し入れ（特に<hp>お菓子</hp>）も大歓迎です
のでよろしくお願いいたします。
<secret>
また、2次会を実施しますので<hp>いつもの
場所</hp>に集まってください。
</body>
```

● DTDファイル

```

:
:
<!ELEMENT hp -- (#PCDATA)*>
<!ENTITY star SDATA "[star]"--=star,open-->
:
:
```



● 正規化パラメタファイル

```
[ELEMENT LIST]
"hp" CONN
"secret" DELT

[CHANGE CHARACTER LIST]
[star] ☆
```



● 正規化結果

```
<body>
<p>
近年、急速に広がりを見せるインターネットに関する
基礎講座を☆〇月×日夕方6時より、本社第一会議室
にて開催いたします。
<note>
なお、閉会後は懇親会を予定しております。差し入れ
（特にお菓子）も大歓迎ですのでよろしくお願いいたします。
</body>
```

2.5.3 SGML 定義情報の登録

Text Search Plug-in で SGML 文書を登録するための定義情報である DTD および正規化パラメタを登録します。

SGML 定義情報は、システム内でユニークな情報として登録名で識別されます。このため、異なる情報は異なる登録名を付けて、システムに登録する必要があります。また、これらの SGML 定義情報の登録名は、CREATE TABLE で列定義時に指定する SGMLTEXT 型にリンクされます。SGML プラグインモジュールが実際に動作する INSERT 処理までに、SGMLTEXT 型データを登録する列に指定した SGML 定義情報

を、データベースに登録する必要があります。

これらの SGML 定義ファイルをデータベースに登録することで、データを登録できるようになります。なお、SGML 定義情報の登録の前に、レジストリを初期化してください。

SGML 定義情報は、ユティリティを使用して次の手順で登録します。

(1) レジストリの初期化

Text Search Plug-in が使用する HiRDB のレジストリの初期化には、「phsregsetup (レジストリ初期化ユティリティ)」を使用します。

レジストリ初期化ユティリティの指定例を次に示します。

```
phsregsetup
```

(2) DTD の登録

DTD の登録には、「phssgmlreg (SGML 定義情報登録ユティリティ)」を使用します。

SGML 定義情報登録ユティリティの指定例を次に示します。

UNIX の場合

```
phssgmlreg DTD MAN.dtd /opt/def1/MANUAL.dtd
```

Windows の場合

```
phssgmlreg DTD MAN.dtd c:¥data¥def1¥MANUAL.dtd
```

(3) 正規化パラメタの登録

正規化パラメタの登録には、「phssgmlreg (SGML 定義情報登録ユティリティ)」を使用します。

SGML 定義情報登録ユティリティの指定例を次に示します。

UNIX の場合

```
phssgmlreg NORparm MANnorm.prm /opt/def1/MANUALnorm.prm
```

Windows の場合

```
phssgmlreg NORparm MANnorm.prm c:¥data¥def1¥MANUALnorm.prm
```

2.5.4 データ格納領域の確保

SGMLTEXT 型を含む表データの格納領域、SGMLTEXT データ本体格納領域、および n-gram インデクス情報格納領域を確保します。データ格納領域の確保には、HiRDB の「pdinit (データベース初期設定ユティリティ)」または「pdmod (データベース構成変更ユティリティ)」の create rdarea 文を使用します。Text Search Plug-in では、次の 3 種類の領域を確保する必要があります。

- SGMLTEXT 型を含む表の基本データ格納領域

2. Text Search Plug-in システムの構築

- SGMLTEXT データ本体格納領域
- n-gram インデクス情報格納領域

SGMLTEXT 型を含む表の基本データ格納領域にはユーザ用 RD エリアを、SGMLTEXT データ本体格納領域と n-gram インデクス情報格納領域にはユーザ LOB 用 RD エリアを確保してください。

create rdarea 文の指定例を図 2-6 および図 2-7 に示します。

図 2-6 create rdarea 文の例 (UNIX の場合)

```
/* テーブル基本データ格納用RDエリア */
create rdarea PDUSER01 for user used by public /* RDエリア名をPDUSER01に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "/DB/DB01/LOB11"
initial 500 segments ;

/* SGMLTEXTデータ格納用RDエリア */
create rdarea PDUSER02 for LOB used by public /* RDエリア名をPDUSER02に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "/DB/DB01/LOB12"
initial 1500 segments ;

/* n-gramインデクス情報格納用RDエリア */
create rdarea PDUSER03 for LOB used by public /* RDエリア名をPDUSER03に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "/DB/DB01/LOB13"
initial 5000 segments ;
```


図 2-7 create rdarea 文の例 (Windows の場合)

```

/* テーブル基本データ格納用RDエリア */
create rdarea PDUSER01 for user used by public /* RDエリア名をPDUSER01に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "c:\%DB%\DB01\LOB11"
initial 500 segments ;

/* SGMLTEXTデータ格納用RDエリア */
create rdarea PDUSER02 for LOB used by public /* RDエリア名をPDUSER02に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "c:\%DB%\DB01\LOB12"
initial 1500 segments ;

/* n-gramインデクス情報格納用RDエリア */
create rdarea PDUSER03 for LOB used by public /* RDエリア名をPDUSER03に設定 */
server name bes1 /* サーバ名の指定bes1 */
storage control segment 1 pages
file name "c:\%DB%\DB01\LOB13"
initial 5000 segments ;

```

2.5.5 表の作成

表の作成には、定義系 SQL の CREATE TABLE を使用します。CREATE TABLE では、文書を登録する列に対する列定義のデータ型として、「SGMLTEXT」と指定してください。

また、抽象データ型定義内 LOB 格納用 RD エリア指定では、属性名には SGMLTEXT を、LOB 属性格納用 RD エリア名には「2.5.4 データ格納領域の確保」で SGMLTEXT データ本体の格納領域として確保した RD エリアを指定してください。

CREATE TABLE のプラグインオプションには、次のフォーマットで登録文書種別および各 SGML 定義情報を指定してください。

CREATE TABLE でのプラグインオプションの形式

```

PLUGIN
' [ 文書登録種別 ]
[ SGML 定義情報 ] ...'

```

オペランド

- 文書登録種別
SGMLTEXT 型の列に登録するテキストデータの種別を指定します。テキストデータの種別を、次のように <TEXTTYPE> タグで囲んだ形式で指定してください。なお、文書登録種別の指定を省略した場合は、SGML 文書として扱います。
- SGML 定義情報
次の内容を指定します。

2. Text Search Plug-in システムの構築

- DTD の登録名称
DTD の登録名称は、<DTD> から始まり、</DTD> で終了する部分に記述します。
- 正規化パラメタの登録名称
正規化パラメタの登録名称は、<NORparm> から始まり、</NORparm> で終了する部分に記述します。
- 文字コード指定（文字コードが utf-8 の場合だけ）
文字コード種別は、<Unicode> から始まり、</Unicode> で終了する部分に「MS」または「JIS」を記述します。指定を省略した場合は、MS が仮定されます。
MS : MS-Unicode
JIS : JISX0221
なお、extracts 関数への引数は MS-Unicode で指定する必要があるため、extracts 関数を使用する場合は「MS」を指定してください。

Text Search Plug-in の使用する文字コードによる、SGML 定義情報の指定の要否を表 2-6 および表 2-7 に示します。

表 2-6 SGML 定義情報の指定（文字コードが Shift-JIS または EUC の場合）

| テキストデータの種別 | 文書登録種別 | DTD 登録名称 | 正規化パラメタ登録名称 |
|------------|----------------------------|----------|-------------|
| SGML 文書 | <TEXTTYPE>SGML</TEXTTYPE> | 必須 | 正規化する場合 |
| XML 文書 | <TEXTTYPE>SGML</TEXTTYPE> | 必須 | 不要 |
| プレーン文書 | <TEXTTYPE>PLAIN</TEXTTYPE> | 不要 | 不要 |

表 2-7 SGML 定義情報の指定（文字コードが utf-8 の場合）

| テキストデータの種別 | 文書登録種別 | DTD 登録名称 | 正規化パラメタ登録名称 | 文字コード指定 |
|------------|----------------------------|----------|-------------|---------|
| XML 文書 | <TEXTTYPE>SGML</TEXTTYPE> | 必須 | 不要 | 指定可 |
| プレーン文書 | <TEXTTYPE>PLAIN</TEXTTYPE> | 指定不可 | 指定不可 | 指定可 |

注

0 バイトの DTD を登録します。ファイル名称は ASCII の範囲内で指定してください。

なお、SGML 定義情報の登録名称には、「phssgmlreg (SGML 定義情報登録コティリティ)」で指定した名称を指定してください。なお、CREATE TABLE の詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

CREATE TABLE の指定例を図 2-8、図 2-9、図 2-10 に示します。

図 2-8 表の定義例 (SGML 文書または XML 文書の場合)

```

CREATE TABLE reports
(
  title      CHAR(32),           ← title列の定義
  news       DATE,              ← news列の定義
  author     CHAR(32),          ← author列の定義
  doc        SGMLTEXT           ← doc (SGMLTEXT型) 列の定義
  ALLOCATE (sgmltext IN (PDUSER02))
            PLUGIN              ← PLUGIN指定
            '<TEXTTYPE>SGML</TEXTTYPE><DTD>MAN.dtd</DTD>'
            '<NORparm>MANnorm.prm</NORparm>' ← SGML定義情報の指定
) IN (PDUSER01);              ← 表格納RDエリアの指定

```

図 2-9 表の定義例 (プレーン文書の場合)

```

CREATE TABLE reports
(
  title      CHAR(32),           ← title列の定義
  news       DATE,              ← news列の定義
  author     CHAR(32),          ← author列の定義
  doc        SGMLTEXT           ← doc (SGMLTEXT型) 列の定義
  ALLOCATE (sgmltext IN (PDUSER02))
            PLUGIN              ← PLUGIN指定
            '<TEXTTYPE>PLAIN</TEXTTYPE>'     ← SGML定義情報の指定
) IN (PDUSER01);              ← 表格納RDエリアの指定

```

図 2-10 表の定義例 (XML 文書で utf-8 (MS-Unicode) の場合)

```

CREATE TABLE reports
(
  title      CHAR(32),           ← title列の定義
  news       DATE,              ← news列の定義
  author     CHAR(32),          ← author列の定義
  doc        SGMLTEXT           ← doc (SGMLTEXT型) 列の定義
  ALLOCATE (sgmltext IN (PDUSER02))
            PLUGIN              ← PLUGIN指定
            '<TEXTTYPE>SGML</TEXTTYPE><DTD>MAN.dtd</DTD>'
            '<Unicode>MS</Unicode>'         ← SGML定義情報の指定
) IN (PDUSER01);              ← 表格納RDエリアの指定

```

2.5.6 インデクスの定義

文書検索プラグインでのインデクスの作成は、通常のインデクス定義、および差分インデクスを使用した登録性能を向上させるためのインデクス定義があります。また、文書の登録または更新時のインデクス作成方法を指定することもできます。

注意事項

大量の文書が登録されている表に対し、CREATE INDEX 文で n-gram インデクス

2. Text Search Plug-in システムの構築

を作成する場合、タイムアウトが発生する場合があります。これを避けるために、大量の文書のインデックスを作成する CREATE INDEX 文を実行するときは、次に示す設定をしてタイムアウトを無制限にしてから実行してください。

HiRDB のシステム共通定義

```
pd_watch_time=0
```

クライアント環境変数

```
PDCWAITTIME=0
```

また、この場合はログを取得しない運用をお勧めします。ログを取得しない運用については、「5.3.1 登録運用のログ取得方式の設定について」の「(3) ログレスモード」を参照してください。

以降、各定義方法について説明します。

(1) 通常のインデックス定義

全文検索用または概念検索用のインデックス作成情報を定義するには、定義系 SQL の CREATE INDEX を使用します。この定義をすることで、全文検索機能や概念検索機能が使用できるようになります。

CREATE INDEX のインデックス型識別子には NGRAM と指定してください。また、RD エリア名には、「2.5.4 データ格納領域の確保」で n-gram インデックス情報の格納領域として確保した RD エリアを指定してください。n-gram インデックスプラグインで使用するインデックスの RD エリアは、ログ取得モードとなります。CREATE INDEX の詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

なお、CREATE INDEX 実行時に指定するプラグインオプションには複数の定義が指定できます。

(例)

複数の定義を指定する場合は、次のように定義ごとに、(半角コンマ)で区切って指定します。

```
PLUGIN '定義,定義,定義,...'
```

インデックスの定義では、全文検索機能だけを使用する場合と、全文検索機能に加えて、概念検索機能を使用する場合とでは、CREATE INDEX の指定方法が異なります。それぞれのインデックスの定義例を次に示します。

全文検索機能を使用する場合

全文検索機能を使用する場合の、インデックスの定義例を次の図に示します。

図 2-11 全文検索機能使用時のインデクス定義例

| | |
|---------------------------|------------------|
| CREATE INDEX REPORTSindex | ← インデクス識別子の指定 |
| USING TYPE NGRAM | ← インデクス型識別子の指定 |
| ON reports(doc) | ← インデクスを定義する列の指定 |
| IN (PDUSER03); | ← 格納RDエリアの指定 |

全文検索機能に加えて、概念検索機能を使用する場合

全文検索機能に加えて、概念検索機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'CONCEPT_ON'
```

この指定を省略した場合、概念検索用のインデクスが作成されないため、概念検索を実行できません。概念検索機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-12 概念検索機能使用時のインデクス定義例

| | |
|--------------------------|---------------------|
| CREATE INDEX REPORTindex | |
| USING TYPE NGRAM | |
| ON reports(doc) | |
| IN (PDUSER03) | |
| PLUGIN 'CONCEPT_ON'; | ← 概念検索用インデクスを作成する指定 |

(2) インデクスファイル自動拡張機能のインデクス定義

インデクスファイル自動拡張機能を利用すると、インデクスファイル作成時の空き容量を、各論理ファイルの拡張領域として使用できるようになります。これによって RD エリアを有効利用できますので、インデクスファイル自動拡張機能の使用を推奨します。この機能を使用する場合は、プラグインオプションに次の定義が必要です。

```
PLUGIN 'FLEX_AREA'
```

この機能を使用するとインデクスが自動的に拡張するため、インデクスの拡張ユティリティ (phnexpnd) を実行する必要はありません。この機能を利用した場合にインデクスの拡張ユティリティを実行したときは、無視されます。

インデクスファイル自動拡張機能を使用する場合の、インデクス定義例を次の図に示します。

図 2-13 インデクスファイル自動拡張機能使用時のインデクス定義例（全文検索用インデクスの場合）

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'FLEX_AREA';    ← インデクスファイル自動拡張機能使用の指定
```

（3）インデクス作成ルール追加機能のインデクス定義

インデクス作成ルール追加機能とは、標準で作成するインデクス情報に、かたかな、漢字、数字、英字など、すべての文字の種類（文字種）の組み合わせを追加する機能です。この機能を使用することで、検索タームがさまざまな文字種の組み合わせで構成される場合の検索性能が向上します。

特に、概念検索機能を使用する場合には、インデクス作成ルール追加機能を使用することをお勧めします。

（a）インデクス作成ルール追加機能の定義方法

インデクス作成ルール追加機能を使用する場合には、CREATE INDEX のプラグインオプションに対して次の表に示す内容を定義してください。

表 2-8 CREATE INDEX のプラグインオプションに対する定義内容

| 定義内容 | 説明 |
|--------------|--|
| INDEX_ALL | すべての文字種の組み合わせを追加します。特に、概念検索機能を使用する場合には、併用することをお勧めします。 |
| INDEX_MEDIUM | かたかな、数字、および記号の一部の文字種の組み合わせを追加します。専門分野で使用する用語や技術系の用語が頻出する場合に、使用することをお勧めします。 |

注

半角の英字で記述します。

インデクス作成ルール追加機能を使用する場合の、インデクスの定義例を図 2-14 および図 2-15 に示します。

図 2-14 インデクス作成ルール追加機能使用時 (INDEX_MEDIUM) のインデクス定義例 (全文検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports (doc)
  IN (PDUSER03)
  PLUGIN 'INDEX_MEDIUM';    ← インデクス作成ルール追加機能使用の指定
```

図 2-15 インデクス作成ルール追加機能使用時 (INDEX_ALL) のインデクス定義例 (概念検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports (doc)
  IN (PDUSER03)
  PLUGIN 'CONCEPT_ON, INDEX_ALL';    ← 概念検索用インデクスと、
                                         インデクス作成ルール追加機能を同時に指定
```

(b) インデクス作成ルール追加機能使用時の注意

インデクス作成ルール追加機能は SGMLTEXT 型の列に対してだけ指定できます。FREEWORD 型の列に対しては指定できません。

(4) 大小文字・全角文字半角文字統一機能のインデクス定義

大文字と小文字, または全角文字と半角文字の違いを区別しないで検索を実行したい場合には, 大小文字・全角文字半角文字統一機能を使用します。

(a) 大小文字・全角文字半角文字統一機能の定義方法

大小文字・全角文字半角文字統一機能を使用する場合には, CREATE INDEX のプラグインオプションに対して次の表に示す内容を定義します。

表 2-9 定義内容

| 定義内容 | 説明 |
|-------------|--|
| SAMECASE=ON | 英大文字と英小文字を区別しないでインデクスを作成します。 |
| SAMEWIDE=ON | 全角文字と半角文字の英数字, かたかな, 記号を区別しないでインデクスを作成します。 |

注

半角の英字で記述します。SAMECASE と SAMEWIDE を組み合わせて指定できます。

大小文字・全角文字半角文字統一機能使用時のインデクス定義例を図 2-16, 図 2-17, 図 2-18 に示します。

2. Text Search Plug-in システムの構築

図 2-16 大小文字統一機能使用時のインデクス定義例（全文検索用のインデクスの場合）

```
CREATE INDEX REPORTindex
USING TYPE NGRAM
ON reports(doc)
IN (PDUSER03)
PLUGIN 'SAMECASE=ON'; ← 大小文字統一機能使用の指定
```

図 2-17 全角文字半角文字統一機能使用時のインデクス定義例（全文検索用のインデクスの場合）

```
CREATE INDEX REPORTindex
USING TYPE NGRAM
ON reports(doc)
IN (PDUSER03)
PLUGIN 'SAMEWIDE=ON'; ← 全角文字半角文字統一機能使用の指定
```

図 2-18 大小文字統一機能，および全角文字半角文字統一機能使用時のインデクス定義例（概念検索用のインデクスの場合）

```
CREATE INDEX REPORTindex
USING TYPE NGRAM
ON reports(doc)
IN (PDUSER03)
PLUGIN 'CONCEPT_ON, SAMECASE=ON, SAMEWIDE=ON'; ← 概念検索用インデクスと、大小文字統一機能と全角文字半角文字統一機能を同時に指定
```

(b) 大小文字・全角文字半角文字統一機能使用時の注意

- 同義語展開指定をする場合、あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか、または同義語辞書移行ユティリティ（「phssyndicconv」）を実行して、既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合、同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。
- extracts 関数でハイライトタグを埋め込む場合、ハイライト条件式に指定する検索タームは文字統一の対象となりません。データを登録した時の文字で指定してください。extracts 関数については、「4.2.6 extracts」を参照してください。
- 全角文字半角文字統一機能を指定しないで、近傍条件検索やワイルドカードに"? "を指定した検索を実行した場合、半角文字のガ行、ザ行、ダ行、バ行、パ行、およびヴを2文字として扱います。しかし、全角文字半角文字統一機能を指定した場合は1文字として扱うため注意が必要です。

(5) 拗音統一機能のインデクス定義

検索時に「や」と「ゃ」、「よ」と「ょ」などの拗音を区別しないようにするには、拗音統一機能を使用します。

(a) 拗音統一機能の定義方法

拗音統一機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義を追加してください。

```
PLUGIN 'SAMEY=ON'
```

拗音統一機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-19 拗音統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports (doc)
  IN (PDUSER03)
  PLUGIN 'SAMEY=ON'; ← 拗音統一機能使用の指定
```

「SAMEY=ON」を指定すると、次の拗音が統一されます。

●全角文字

あ, い, う, え, お, つ, や, ゆ, よ, ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

●半角文字

ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

(b) 拗音統一機能使用時の注意

- 同義語展開指定をする場合、あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか、または同義語辞書移行ユティリティ (「phssyndicconv」) を実行して、既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合、同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。
- extracts 関数でハイライトタグを埋め込む場合、ハイライト条件式に指定する検索タームは文字統一の対象となりません。データを登録した時の文字で指定してください。extracts 関数については、「4.2.6 extracts」を参照してください。

(6) 濁音統一機能のインデクス定義

検索時に「パ」と「パ`」, 「ぱ」や「ぱ`」のように濁音や半濁音を区別しないようにするには、濁音統一機能を使用します。

(a) 濁音統一機能の定義方法

濁音統一機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義を追加してください。

```
PLUGIN 'SAMED=ON'
```

濁音統一機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-20 濁音統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'SAMED=ON';    ← 濁音統一機能使用の指定
```

「SAMED=ON」を指定すると、濁音、半濁音が統一されます。

(b) 濁音統一機能使用時の注意

- 同義語展開指定をする場合、あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか、または同義語辞書移行ユティリティ (「phssyndicconv」) を実行して、既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合、同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。
- extracts 関数でハイライトタグを埋め込む場合、ハイライト条件式に指定する検索タームは文字統一の対象となりません。データを登録した時の文字で指定してください。extracts 関数については、「4.2.6 extracts」を参照してください。
- 濁音文字統一機能を指定しないで、近傍条件検索やワイルドカードに "?" を指定した検索を実行した場合、半角文字の濁音で表示できる文字を 2 文字として扱います。しかし、濁音文字統一機能を指定した場合は 1 文字として扱うため注意が必要です。

(7) 削除文字指定機能のインデクス定義

検索に使用しない文字を除いて検索性能を向上させるには、削除文字指定機能を使用します。

(a) 削除文字指定機能の定義方法

削除文字指定機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'DELcode=ON'
```

削除文字指定機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-21 削除文字指定機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'DELcode=ON';    ← 削除文字指定機能使用の指定
```

「DELcode=ON」を指定すると、次の文字が削除されます。

全角文字の "。(句点), "。(読点), 全角スペース,
 半角文字の "。(句点), "。(読点), 半角スペース (0x20),
 改行コード (0x0a, 0x0d), タブコード (0x09)

(b) 削除文字指定機能使用時の注意

- 近傍検索で指定された語句リストの間に削除文字が含まれる場合、距離の指定方法によってはヒットしないことがあります。
 例えば、登録データが「COMPUTER NETWORK」の場合、「COMPUTER」と「NETWORK」の間の距離は0となります。
- 削除文字を指定したインデクスに対して検索を実行する場合、登録データから削除文字が取り除かれていることに注意してください。
 例えば、削除文字が半角スペースで、登録データが「tha[. . .]</終了タグ>」の場合（. . . は0個以上の繰り返しを示します）、「tha?」で検索してもヒットしません。
- 検索タームおよびすべての異表記展開・同義語異表記展開結果が、削除文字だけで構成されている場合、0件ヒットとなります。
- 削除文字を指定した検索結果に対し、extracts 関数でハイライト表示しても、ハイライト表示する文字列に削除文字が含まれている場合、ハイライト表示されません。

(8) 文字単位インデクス作成機能使用時のインデクス定義

インデクス容量を削減する方法として、1文字ずつインデクスを作成することができます。この機能を文字単位インデクス (1gram インデクス) 作成機能といいます。

(a) 文字単位インデクス作成機能の定義方法

文字単位インデクス作成機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'INDEX_GRAM1'
```

文字単位インデクス作成機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-22 文字単位インデクス作成機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'INDEX_GRAM1';    ← 文字単位インデクス作成機能使用の指定
```

(b) 文字単位インデクス作成機能使用時の注意事項

- 文字単位インデクス作成機能は、次の機能と同時に指定することはできません。
 - 概念検索用インデクスの作成

2. Text Search Plug-in システムの構築

- インデクス作成ルール追加機能の使用
- 文字単位インデクスだけ作成した場合、インデクス容量が削減される反面、登録データの増加に伴い、検索性能が低下します。このため、大量データを登録する運用にこの機能を適用する場合は注意が必要です。

(9) 差分インデクス定義

登録性能を向上させるために、小容量の一時的な登録用インデクスに、文書を追加します。この登録用インデクスを差分インデクスといいます。

(a) 差分インデクスの定義方法

差分インデクスを使用する場合は、CREATE INDEX のプラグインオプションに差分インデクスとして使用する容量を定義します。サイズはキロバイト単位で、次のように定義します。

```
PLUGIN 'SUB_INDEX=xxxxxx'
```

なお、差分インデクスとして使用できる容量の最大値は 102,400 キロバイト、最小値は 40,000 キロバイトです。差分インデクスの定義が省略された場合、差分インデクスは作成されません。

差分インデクス使用時のインデクスの定義例を図 2-23 および図 2-24 に示します。

図 2-23 差分インデクス使用時のインデクス定義例 (全文検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'SUB_INDEX=51200'; ← 差分インデクスとして使用する容量の指定
```

図 2-24 差分インデクス使用時のインデクス定義例 (概念検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'CONCEPT_ON, SUB_INDEX=51200'; ← 概念検索用のインデクスと差分インデクスを同時に作成
```

(b) 登録先インデクスの指定方法

差分インデクスを作成した場合、ユーザは HiRDB クライアント環境変数の指定によって、MASTER インデクスへの登録か、差分インデクスへの登録かを選択できます。

HiRDB クライアント環境変数は次のように指定します。

環境変数名称 : PDPLUGINNSUB

指定できる値 Y または N

Y : 差分インデクスへの登録

N : MASTER インデクスへの登録

環境変数に指定がない場合は、"Y" (差分インデクスへの登録) が仮定されます。ただし、次の場合には環境変数の指定内容に関係なく、MASTER インデクスへ登録します。

- 差分インデクスを作成していない場合
- インデクス作成後またはインデクス再作成後に、インデクスへの文書登録が 0 件の場合
- 差分インデクス使用率が 100% になった場合

(c) 差分インデクス使用時の注意

- HiRDB の「pload (データベース作成ユティリティ)」および「pdrorg (データベース再編成ユティリティ)」によるインデクス登録時も、差分インデクスへの登録ができます。この場合は、作成するインデクス量が差分インデクス残容量を超えないように注意してください。残容量を超えた場合は、システムログファイルに次のメッセージが出力されます。

```
KFPY99999-I (_phngram) 02000-I Sub index using ratio 100 percent
```

この場合、MASTER インデクスへの登録が完了すれば、pload および pdrorg は正常終了します。ただし、以降の登録および更新処理で差分インデクスを使えるようにするために、差分インデクスのマージをしてください。

- HiRDB の「pload (データベース作成ユティリティ)」を使用してデータ登録を実行する場合に、環境変数の指定をしていないと、差分インデクスへの登録が実行されず。この時に差分インデクス容量が不足すると、MASTER インデクスに再度登録が実行されることになり、余計に時間が掛かります。このため、環境変数の設定値を確認してから、「pload (データベース作成ユティリティ)」を実行してください。
- UCS4 インデクスでは指定できません。

(10) インデクス作成方法指定

表の SGMLTEXT 型列に文書を登録している状態で CREATE INDEX を実行する場合、次のどれかの方法を指定できます。運用形態に合わせて指定してください。

CREATE INDEX 実行時にインデクス作成とインデクスへの文書登録を実行する方法 (即時登録)

CREATE INDEX 実行時にはインデクスに反映しないで文書やデータの登録だけを実行し、「phnstartidx (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の文書登録を実行する方法 (初期分割登録または分割遅延登録)

(a) インデクス作成方法指定の定義方法

インデクス作成方法指定を使用する場合、CREATE INDEX のプラグインオプションに即時登録、初期分割登録、または分割遅延登録の定義を追加します。インデクス作成方法指定の定義を省略した場合、即時登録が適用されます。

2. Text Search Plug-in システムの構築

即時登録

表に登録されている文書について、CREATE INDEX 実行時にインデクス作成とインデクスへの文書登録を実行します。

即時登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加、または省略します。

```
PLUGIN 'CREATEMODE=0'
```

初期分割登録

表に登録されている文書について、CREATE INDEX 実行時にはインデクスへは反映しません。「phnstartidx (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の文書登録を実行します。

初期分割登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=1'
```

初期分割登録使用時のインデクスの定義例を図 2-25 および図 2-26 に示します。

図 2-25 初期分割登録使用時のインデクス定義例 (全文検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports (doc)
  IN (PDUSER03)
  PLUGIN 'CREATEMODE=1'; ← 初期分割登録の指定
```

図 2-26 初期分割登録使用時のインデクス定義例 (概念検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports (doc)
  IN (PDUSER03)
  PLUGIN 'CONCEPT_ON, CREATEMODE=1'; ← 概念検索用のインデクスと
                                         初期分割登録を同時に指定
```

分割遅延登録

表に登録されている文書について、CREATE INDEX 実行時にはインデクスへ反映しません。「phnstartidx (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の文書登録を実行します。分割遅延登録を使用する場合、インデクス分割遅延登録用の領域が必要となります。

分割遅延登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=2, DELAY_SIZE=インデクス分割遅延登録用領域サイズ
(キロバイト)'
```

インデクス分割遅延登録用領域サイズの見積もりについては、「2.3.2(1)(b) ユーザ

「LOB 用 RD エリア容量」を参照してください。また、分割遅延登録を設定していないインデクスに対しては、「phnmodidx (インデクス情報変更ユティリティ)」を実行することで分割遅延登録を設定したインデクスに移行できます。指定できる最大値・最小値は次のとおりです。

- 最大値：129,584 キロバイト
- 最小値：2,608 キロバイト

分割遅延登録時のインデクスの定義例を図 2-27 および図 2-28 に示します。

図 2-27 分割遅延登録使用時のインデクス定義例 (全文検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'CREATEMODE=2, DELAY_SIZE=10240'; ← 分割遅延登録の指定, インデクス分割遅延登録用領域サイズを10,240(キロバイト)に設定
```

図 2-28 分割遅延登録使用時のインデクス定義例 (概念検索用のインデクスの場合)

```
CREATE INDEX REPORTindex
  USING TYPE NGRAM
  ON reports(doc)
  IN (PDUSER03)
  PLUGIN 'CONCEPT_ON, CREATEMODE=2, DELAY_SIZE=10240'; ← 概念検索用のインデクスと分割遅延登録の指定, インデクス分割遅延登録用領域サイズ領域サイズを10,240(キロバイト)に設定
```

分割遅延登録への移行時に実行する、「phnmodidx (インデクス情報変更ユティリティ)」の指定例を次の図に示します。

図 2-29 分割遅延登録への移行時に実行する、インデクス情報変更ユティリティの指定例

```
phnmodidx -d REPORTindex -v "DELAY_SIZE=10240"; ← インデクス分割遅延登録用領域サイズを10,240(キロバイト)に設定
```

分割遅延登録・更新系 SQL 同時実行

遅延登録実行時に、同時に「phnstartidx (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を登録します。分割遅延登録・更新系 SQL 同時実行を利用する場合、遅延登録時に情報を退避しておく領域が必要です。CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=3, DELAY_SIZE=遅延登録時に情報を退避する領域サイズ(キロバイト)'
```

2. Text Search Plug-in システムの構築

遅延登録時に情報を退避する領域サイズの見積もりについては、「2.3.2(1)(b) ユーザ LOB 用 RD エリア容量」を参照してください。

この退避領域に指定できる最大値・最小値は次のとおりです。

- 最大値：129,584 キロバイト
- 最小値：12,832 キロバイト

分割遅延登録・更新系 SQL 同時実行のインデクスの定義例を次の図に示します。

図 2-30 分割遅延登録・更新系 SQL 同時実行使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(doc)
  IN (PDUUSER03)
  PLUGIN 'CREATEMODE=3, DELAY_SIZE=12832'; ← 分割遅延登録・インデクス作成同時実行の指定,
                                              遅延登録時に情報を退避させるスタック領域サ
                                              イズを12,832(キロバイト)に設定
```

(b) 初期分割登録・分割遅延登録使用時の注意

- CREATE INDEX 実行後、インデクスへ登録されていない文書のインデクスは、「phnstartidx (インデクス作成開始ユティリティ)」を実行して登録してください。インデクスへ登録されていない文書が存在するかどうかは、「phnidxls (インデクス情報の取得ユティリティ)」を実行して確認してください。
- インデクスへ登録されていない文書は、検索対象になりません。
- インデクスへ登録されていない文書が存在する場合、登録または更新した文書はインデクスへ登録されません。インデクスへ登録する場合は、「phnstartidx (インデクス作成開始ユティリティ)」を実行してください。
- インデクスが作成されていない文書が存在する場合、検索を実行しても、インデクスが作成されていない旨のメッセージは出力されません。
- 差分インデクスを使用している場合、登録または更新した文書はすべて MASTER インデクスに登録されます。
- 初期分割登録の場合、CREATE INDEX 実行後、インデクス作成開始ユティリティを実行して、すべての文書がインデクスへ登録された時点で、インデクス作成方法が初期分割登録から即時登録に切り替わります。以降、文書の登録または更新と同時にインデクスへの登録が行われるようになるため、文書登録または更新時にインデクスへ登録する時間が掛かるようになります。
- 「phnmodidx (インデクス情報変更ユティリティ)」で移行したインデクスに対して初期化をした場合、変更した情報は無効になります。
- インデクス分割遅延登録用領域を使い切った場合は、遅延しているすべての文書をインデクスに反映するか、または「phnmodidx (インデクス情報変更ユティリティ)」を使用してインデクス分割遅延登録用領域を拡張してください。
- 分割遅延登録・更新系 SQL 同時実行をしない運用へ移行する場合、インデクスを削除してください。そのあと、運用に合わせてインデクスを再定義してください。

(11) UCS4 のインデクス定義

UCS2 の範囲を超える UCS4 の範囲の文字を使用するには、UCS4 用の定義が必要になります。また、UCS4 インデクスは自動拡張機能を適用します。

(a) インデクスの定義方法

UCS2 の範囲を超える UCS4 の範囲の文字を使用する場合は、プラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'INDEX_UCS4'
```

(b) インデクス定義時の注意

- UCS4 用のインデクス定義の追加をしていない状態で UCS2 の範囲を超える文字を登録した場合は、該当文字は無視されます。
- UCS4 用のインデクス定義の追加をしていない状態で UCS2 の範囲を超える文字を検索した場合は、エラーとなります。
- HiRDB の文字コード種別が utf-8 以外の時に INDEX_UCS4 を指定した場合は、インデクス定義はエラーとなります。
- 以下のプラグインオプションは INDEX_UCS4 と同時に指定することができません。
 - 差分インデクス (SUB_INDEX)
 - 英文インデクス (ENGLISH, ENGLISH_STANDARD)

(c) UCS2 から UCS4 への移行

UCS2 と UCS4 はインデクスの構成が異なるため、UCS2 用のインデクスをそのまま UCS4 用のインデクスとして使用することはできません。

UCS2 用のインデクスを UCS4 用として使用する場合は、「`phnmodidx` (インデクス情報変更ユティリティ)」で変更項目に "INDEX_TYPE=UCS4" を指定して移行してください。

移行する前のインデクスに次の機能が指定されている場合は、移行コマンド実行時にエラーとなります。

- 差分インデクス (SUB_INDEX)
- 英文インデクス (ENGLISH, ENGLISH_STANDARD)
- データベースの文字コード種別が utf-8 ではないデータベースの環境

(12) 英文検索機能のインデクス定義

英文に対する検索を行なう場合は、単語単位での検索を行なう英文インデクス定義をしてください。

英文インデクスの定義は次の方法があります。運用に応じてご使用ください。

- 単語一致 (ENGLISH)

2. Text Search Plug-in システムの構築

- 単語一致，前方一致（ENGLISH_STANDARD）

図 2-31 英文検索機能を使用する場合のインデクス定義例（英文検索）

| | |
|--|--|
| <pre>CREATE INDEX REPORTindex USING TYPE NGRAM ON reports(text) IN (PDUUSER03) PLUGIN 'ENGLISH_STANDARD, INDEXd1mt=INDEXd1mt.prm, NOindex=NOindex.prm, DELcode=DELcode.prm';</pre> | <ul style="list-style-type: none">← インデクス識別子の指定← インデクス種別の指定← インデクスを定義する列の指定← 格納 R D エリアの指定← 英文検索用インデクスを作成する指定, インデクスデリミタファイル名の指定, 不要語ファイル名の指定, 削除コードファイル名の指定 |
|--|--|

注意事項

英文インデクス定義は，次のインデクス定義と併用することはできません。

- 概念検索機能（CONCEPT_ON）
- 全角文字半角文字統一機能（SAMEWIDE=ON）
- 拗音統一検索（SAMEY=ON）
- インデクス作成ルール追加機能（INDEX_ALL）
- インデクス作成ルール追加機能（INDEX_MEDIUM）
- 文字単位インデクス作成機能（INDEX_GRAM1）

(a) 英文検索のインデクスに指定できる項目

英文検索機能で使用するインデクスの定義ファイルについて説明します。

インデクスデリミタファイル

機能

英文検索機能で使用するインデクスを作成する場合に必要な単語の区切り（デリミタ）を定義します。定義したインデクスデリミタに従って単語を抽出し，英文検索機能で使用するインデクスを作成します。抽出した単語が 64 文字（バイト）を超える場合は，65 文字（バイト）目以降を無視して，64 文字（バイト）目までの文字でインデクスを作成します。なお，インデクスデリミタは検索対象にできません。

指定形式

CREATE INDEX 文のプラグインオプションにインデクスデリミタファイルのファイル名を指定します。

INDEXd1mt=ファイル名

インデクスデリミタファイルは，CREATE INDEX を実行する前に，Text Search Plug-in をセットアップしたすべてのサーバマシンに対して，同一の内容で次に示すディレクトリに事前に作成しておく必要があります。

\$PDDIR/plugin/_phngram/idx_def

また、サンプルファイル（\$PDDIR/plugin/_phngram/sample/pdplgdlmt）を提供しています。

記述形式

"インデクスデリミタ" [,"インデクスデリミタ"]・・・・

- インデクスデリミタは,"(半角ダブルクォーテーション)で囲んで指定します。
- インデクスデリミタに使用できる文字は,'0x21' ~ '0x7E'の範囲の1バイト文字コードです。
- '0x20'(半角スペース), '0x09'(タブ)は,無条件にインデクスデリミタとなります。
- 複数指定する場合は,"(半角ダブルクォーテーション)で囲んだインデクスデリミタを,(半角コンマ)で区切って指定します。
- "(半角ダブルクォーテーション)をインデクスデリミタにする場合は,""と二つ続けて定義し,それを"(半角ダブルクォーテーション)で囲んで指定します。

注意事項

- インデクスデリミタファイルは省略できます。
- 省略した場合は,'0x20'(半角スペース), '0x09'(タブ), 記号「'0x21' ~ '0x2F', '0x3A' ~ '0x40', '0x5B' ~ '0x60', '0x7B' ~ '0x7E'」がインデクスデリミタとなります。
- 不要語,または削除コードに指定した文字コードをインデクスデリミタに指定しないでください。指定した場合は,インデクスデリミタ,削除コード,不要語,の優先順位で処理します。

不要語ファイル

機能

英文検索機能に使用するインデクスを作成する場合,検索に不要な単語を定義します。不要語については,検索用インデクスを作成しません。

指定形式

CREATE INDEX 文のプラグインオプションに不要語ファイルのファイル名を指定します。

NOindex=ファイル名

不要語ファイルは,CREATE INDEX を実行する前に,Text Search Plug-in をセットアップしたすべてのサーバマシンに対して,同一の内容で次のディレクトリに事前に作成しておく必要があります。

\$PDDIR/plugin/_phngram/idx_def

記述形式

" 不要語 " [," 不要語 "]

- 不要語は, " (半角ダブルクォーテーション) で囲んで指定します。
- 不要語に使用できる文字は, '0x21' ~ '0x7E' の範囲 1 バイトの文字コードです。
- 複数指定する場合は, " (半角ダブルクォーテーション) で囲んだ不要語を, (半角コンマ) で区切って指定します。不要語は 100 個まで指定できます。
- " (半角ダブルクォーテーション) を不要語にする場合は, "" と二つ続けて定義し, それを " (半角ダブルクォーテーション) で囲んで指定します。
- 不要語は 64 文字 (バイト) まで指定できます。64 文字 (バイト) を超えて指定した場合, 65 文字 (バイト) 目以降を無視して, 64 文字 (バイト) 目までの文字列を不要語とします。

注意事項

- 不要語ファイルは省略できます。省略した場合は, 不要語なしでインデクスを作成します。
- インデクスデリミタ, または削除コードに指定した文字コードを不要語に指定しないようにしてください。指定した場合は, インデクスデリミタ, 削除コード, 不要語の優先順位で処理します。
- 検索条件で指定した検索タームに不要語を含む場合は, 不要語を任意の 1 単語とみなして検索します。なお, 検索条件に不要語だけを指定した場合は, 入力条件エラーとなります。

削除コードファイル

機能

英文検索機能で使用するインデクスを作成する場合に, 削除するコードを定義します。定義した削除コードはないものとして単語を抽出し, 英文検索機能で使用するインデクスを作成します。

単語に - (ハイフン) のような不要なコードがある場合, 削除してその前後を詰めて扱い, 必要なコードだけでインデクスを作成できます。

指定形式

CREATE INDEX 文のプラグインオプションに削除コードファイルのファイル名を指定します。

DELcode=ファイル名

削除コードファイルは, CREATE INDEX を実行する前に, Text Search Plug-in をセットアップしたすべてのサーバマシンに対して, 同じ内容で次のディレクトリに事前に作成しておく必要があります。

\$PDDIR/plugin/_phngram/idx_def

記述形式

"削除コード" [,"削除コード"].....

- 削除コードは," (半角ダブルクォーテーション) で囲んで指定します。
- 削除コードに使用できる文字は,'0x21' ~ '0x7E' の範囲の 1 バイト文字コードです。
- 複数指定する場合は," (半角ダブルクォーテーション) で囲んだ削除コードを,(半角コンマ) で区切って指定します。
- " (半角ダブルクォーテーション) を削除コードにする場合は,"" と二つ続けて定義し,それを" (半角ダブルクォーテーション) で囲んで指定します。

注意事項

- 削除コードファイルは省略できます。省略した場合は,削除コードなしでインデクスを作成します。
- インデクスデリミタ,または不要語に指定した文字コードを削除コードに指定しないようにしてください。指定した場合は,インデクスデリミタ,削除コード,不要語,の優先順位で処理します。

2.5.7 データの登録

(1) データの登録方法

作成した表にデータを登録するには,操作系 SQL の INSERT 文を使用します。登録するテキストデータは,HiRDB をセットアップしたときの文字コード,および「付録 A 登録する文書 (テキストデータ) および文字列データに使用できる文字」に示す文字コードを使用してください。INSERT 文の詳細については,マニュアル「HiRDB SQL リファレンス」を参照してください。

INSERT 文の指定例を次の図に示します。

図 2-32 INSERT 文の処理の例

```

EXEC SQL BEGIN DECLARE SECTION;
SQL TYPE IS BLOB(1M) sgml;           ← BLOB型変数の定義
EXEC SQL END DECLARE SECTION;

strcpy(sgml.sgml_data,              ← BLOB型変数への代入
char_ptr_pointing_to_a_sgml_text);

sgml.sgml_length =                  ← BLOB型変数データ長の代入
strlen(char_ptr_pointing_to_a_sgml_text);

EXEC SQL                             ← INSERT文の実行
INSERT INTO reports(doc) VALUES(SGMLTEXT(:sgml AS BLOB(1M)));

```

また,HiRDB の「pload (データベース作成ユーティリティ)」を使用すると,データの一括登録ができます。

なお,SQL による SGML 文書登録時に SGML 文書のエラーを検知しても,トランザク

2. Text Search Plug-in システムの構築

ションはロールバックしません。SGML 文書のエラーが発生した場合、SQL (INSERT 文または UPDATE 文) による登録のときは、SQLCODE = -997 のエラーとなります。なお、エラーコード詳細化指定をしている場合は、SQLCODE = -3040 のエラーとなります。

また、pload による登録の場合は、該当する行の登録がスキップされます。

また、文書検索プラグインを使用して SGML 文書を登録する場合、HiRDB のユーザ LOB 用 RD エリアにデータを登録する場合よりも時間が掛かります。これは、登録時に、SGML 文書の構造解析処理を実行して、全文検索をするための文書構造インデクスを作成するためです。登録に掛かる時間は、文書長、DTD の内容、および文書構造の複雑さに依存して変化します。

(2) 登録・検索同時実行の設定

登録・検索同時実行を設定すると、データの登録と検索の同時実行ができます。同時実行には次の利点があります。

- ほかのユーザが更新中 (トランザクションが終了していない状態) の文書を検索できます。
- SGMLTEXT データ格納用 RD エリアおよび n-gram インデクス情報格納用 RD エリアに対してグローバルバッファを割り当てると、登録および検索性能が向上します。

ただし、同時実行できるのは次の場合です。

- SQL による単数の更新系トランザクションと、複数の検索系トランザクションの同時実行
- インデクス未作成の場合、複数の更新系トランザクションの同時実行
- 「phnidxls (インデクス情報の取得ユティリティ)」、「phnmerge (インデクスのマージユティリティ)」、「phnstartidx (インデクス作成開始ユティリティ)」、または「phnconget (統計情報収集ユティリティ)」と、検索系トランザクションの同時実行

(a) 指定方法

Text Search Plug-in の環境設定項目、および HiRDB のバックエンドサーバ定義またはシングルサーバ定義に、次の内容を設定します。Text Search Plug-in の環境設定項目については、「2.6 文書検索プラグインのユーザ環境設定」を参照してください。また、HiRDB のバックエンドサーバ定義またはシングルサーバ定義の詳細については、マニュアル「HiRDB システム定義」を参照してください。

文書検索プラグインの環境設定項目

- SGML プラグインを使用している場合
SGML プラグインの環境設定項目「登録・検索同時実行指定」で次のように指定します。

```
set phs_search_wait=nowait
```

- n-gram インデクスプラグインを使用している場合
n-gram インデクスプラグインの環境設定項目「1RD エリア当たりの排他制御用 共用メモリサイズ」で次のように指定します。

```
set phn_lock_pool_size=共用メモリサイズ
```

HiRDB のバックエンドサーバ定義またはシングルサーバ定義

- n-gram インデクスプラグインを使用している場合
n-gram インデクスプラグインで使用する共用メモリのサイズを指定します。
HiRDB のバックエンドサーバ定義またはシングルサーバ定義のオペランド「pdplgprm」は、次の形式で指定してください。

オペランドの指定形式

```
pdplgprm -n プラグイン名 -s 共用メモリサイズ
```

引数

-n プラグイン名

共用メモリを使用するプラグインの名称を指定します。ここでは、n-gram インデクスプラグインのプラグイン名「_phngram」を指定してください。

-s 共用メモリサイズ

n-gram インデクスプラグインで使用する共用メモリのサイズ（キロバイト）を指定します。

(b) 注意事項

- HiRDB が提供しているユティリティによる登録と検索系トランザクションは同時に実行できません。
- 「phnidxls（インデクス情報の取得ユティリティ）」、「phnmerge（インデクスのマージユティリティ）」、「phnstartidx（インデクス作成開始ユティリティ）」、および「phnconget（統計情報収集ユティリティ）」以外の Text Search Plug-in のユティリティと、検索系トランザクションは同時に実行できません。
- 同時実行を設定している場合、Text Search Plug-in のデータの整合性確保のために、一時的に排他制御をします。このため、同時実行を設定していない場合に比べて、排他制御によるオーバーヘッドが掛かります。
- "phs_search_wait=nowait" を設定した場合は、n-gram インデクスを作成していない場合でも共用メモリの定義をする必要があります。
- 同時実行する場合は、グローバルバッファを指定してください。実行性能が向上します。
- 登録・検索同時実行時に、更新系のトランザクションを複数実行した場合、排他によって wait 状態になります。更新系トランザクションの wait 状態が解消されるまで、以降の検索処理も wait 状態になります。
- LOCK 文でテーブル全体を排他すると、ほかのトランザクションとの同時実行ができません。このため、分割遅延登録・更新系 SQL 同時実行を実施する場合は、LOCK 文でテーブル全体の排他を実施しないようにしてください。

(3) 分割遅延登録・更新系 SQL 同時実行の設定

分割遅延登録・更新系 SQL 同時実行を利用すると、(2) で説明した登録・検索同時実行の機能に加えて、さらに、データ登録とインデクス作成を同時に実行できます。

(a) 指定方法

SGML プラグインの環境設定項目「登録・検索同時実行」に、次のように指定します。

```
set phs_search_wait=nowait_with_update
```

(b) 注意事項

- LOCK 文でテーブル全体を排他すると、ほかのトランザクションとの同時実行ができません。このため、分割遅延登録・更新系 SQL 同時実行を実施する場合は、LOCK 文でテーブル全体の排他を実施しないようにしてください。
- 検索トランザクションの排他が更新系 SQL または phnstartidx (インデクス作成開始ユーティリティ) と競合することを避けるため、検索 SQL 実行時は、SQL の排他オプションに WITHOUT LOCK NOWAIT を指定してください。

2.5.8 同義語辞書の作成

全文検索で検索タームの同義語を検索条件として使用する場合、または概念検索で種文章から抽出した検索用特徴タームの同義語を検索条件として使用する場合は同義語辞書の作成が必要です。

同義語辞書を作成する前に、同義語辞書の基になるファイルを作成します。このファイルを同義語辞書テキストファイルといいます。同義語辞書テキストファイルには、キーワードとなる単語 (検索ターム) の同義語の展開方法を定義します。同義語辞書テキストファイルはテキストエディタなどで作成します。なお、同義語辞書を作成しない場合、このファイルを作成する必要はありません。

(1) 同義語辞書テキストファイルの記述形式

同義語辞書テキストファイルには次の形式で同義語を定義します。

記述形式

同義語, 同義語 [, 同義語] ...

注意事項

- 同義語辞書は、HiRDB をセットアップしたときの文字コード、および「付録 A 登録する文書 (テキストデータ) および文字列データに使用できる文字」に示す文字コードで記述してください。
- 一つの同義語のグループには、2 語以上の同義語を記述してください。
- 行の先頭文字に * (半角アスタリスク) を記述すると、注釈行とみなされ、同義語として扱われません。"*" を同義語とする場合、"¥*" と記述してください。なお、"¥" は半角でも全角でもかまいません。
- 行の途中に (半角スペース) があつた場合、半角スペース以降のデータは注釈

として扱われます。ただし、行の先頭にある半角スペースは無視されます。

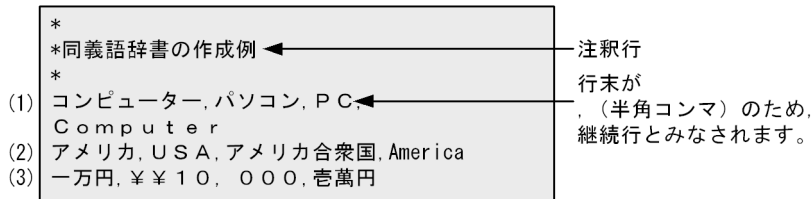
半角スペースを同義語とする場合、"¥ "と記述してください。なお、"¥"は半角でも全角でもかまいません。

- 行の末尾が、(半角コンマ)で終了している場合、行が継続しているものとして扱われます。","を同義語とする場合、"¥,"と記述してください。なお、"¥"は半角でも全角でもかまいません。
- "¥" (エスケープ文字)を同義語とする場合、"¥¥"と記述してください。なお、"¥"は半角でも全角でもかまいません。

(2) 同義語辞書テキストファイルの記述例

同義語辞書テキストファイルの記述例を次の図に示します。

図 2-33 同義語辞書テキストファイルの記述例



(1), (2), (3)の単位で指定されたそれぞれの単語が互いに同義語として扱われます。

例えば、(1)では、「コンピューター」「パソコン」「PC」「Computer」がそれぞれ互いに同義語の関係にあります。

(3) 同義語グループ内での文字種の制限

一つの同義語のグループには、かたかな、漢字、数字、アルファベットなどの種類の文字が混在している場合があります。これらの文字の種類を、文字種といいます。例えば、「アメリカ合衆国」という単語は、かたかな(アメリカ)と漢字(合衆国)の二つの文字種で構成されています。

同義語辞書テキストファイルを作成する場合、次のことに注意してください。

- 一つの同義語のグループを文字種ごとに分割したとき、その数の合計が1,000以下となるようにしてください。ただし、ひらがなは文字種として数えません。
- 長音およびハイフンは、直前にある文字と同じ文字種とします。直前に文字がない場合は、そのほかの扱いになります。

同義語のグループでの文字種の分割例を次の図に示します。

図 2-34 同義語のグループでの文字種の分割例

同義語のグループ

旧 ソビエト 連邦、 ロシア 共和国
1 2 3 4 5 → 分割数の合計 : 5

注 この例では、わかりやすさを強調するため、半角スペースを入れています。実際に、同義語を定義する場合、行の途中に半角スペースが入ると、それ以降のデータは注釈とみなされます。

2.5.9 同義語辞書の登録

同義語辞書の登録には、「phssyndicreg (同義語辞書登録ユーティリティ)」を使用します。

同義語辞書登録ユーティリティの指定例を次に示します。

UNIX の場合

```
phssyndicreg USERDIC01 /opt/def1/mydictionary.dic
```

Windows の場合

```
phssyndicreg USERDIC01 c:¥data¥def1¥mydictionary.dic
```

注意事項

- 文字列検索プラグインで登録した同義語辞書は、文書検索プラグインでは使用できません。文字列検索プラグインと同じ同義語辞書を使用する場合は、文字列検索プラグインで登録した同義語辞書テキストファイルを、文書検索プラグインにも同様に登録してください。

2.5.10 SGML 定義情報のメンテナンス

SGML 文書の構造変更などによって、DTD や正規化パラメタを変更・削除する場合の操作について説明します。

(1) SGML 定義情報の変更

DTD、正規化パラメタを変更するには、「phssgmlrm (SGML 定義情報削除ユーティリティ)」で定義情報を削除したあと、「phssgmlreg (SGML 定義情報登録ユーティリティ)」で変更後の情報を登録します。SGML 定義情報登録ユーティリティで指定する登録名には、変更前の情報を登録した名称と同じ名称を指定してください。

DTD ファイルおよび正規化パラメタファイルの変更は、すでに登録したデータに影響する場合がありますので注意が必要です。表にデータが登録済みで変更をする場合は、表に登録済みのデータをいったん削除し、DTD または正規化パラメタを変更したあと、データを再登録してください。

SGML 定義情報削除ユーティリティおよび SGML 定義情報登録ユーティリティの指定例を次

に示します。

UNIX の場合

```
phssgmlrm DTD MAN.dtd
phssgmlreg DTD MAN.dtd /opt/def2/MANUAL.dtd
```

Windows の場合

```
phssgmlrm DTD MAN.dtd
phssgmlreg DTD MAN.dtd c:¥data¥def2¥MANUAL.dtd
```

(2) SGML 定義情報の削除

DTD, 正規化パラメタの削除には, 「phssgmlrm (SGML 定義情報の削除ユティリティ)」を使用します。SGML 定義情報の削除ユティリティで指定する登録名には, 削除する情報を登録した名称と同じ名称を指定してください。

SGML 定義情報削除ユティリティの指定例を次に示します。

```
phssgmlrm NORparm MANnorm.prm
```

(3) SGML 定義情報の参照

DTD, 正規化パラメタを参照するには, 「phssgmlval (SGML 定義情報参照ユティリティ)」を使用します。SGML 定義情報参照ユティリティで指定する登録名には, 参照したい情報を登録した名称と同じ名称を指定してください。

SGML 定義情報参照ユティリティの指定例を次に示します。

UNIX の場合

```
phssgmlval NORparm MANnorm.prm /tmp/normparm.txt
```

Windows の場合

```
phssgmlval NORparm MANnorm.prm c:¥temp¥normparm.txt
```

2.6 文書検索プラグインのユーザ環境設定

この節では、Text Search Plug-in システムの文書検索プラグインの環境をユーザごとに設定する方法について説明します。

2.6.1 SGML プラグインの環境設定

(1) 環境情報の定義方法

SGML プラグインの環境情報を設定するファイルを環境定義ファイルといいます。環境定義ファイルは、プラグインごとに作成します。SGML プラグインの環境定義ファイルは、サンプルファイルとして提供されています。

SGML プラグインの環境情報を設定できる単位を次に示します。

システム単位

システム単位に定義情報を設定するユーザは、プラグインシステム定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phsgml/conf/pdplgsys
```

- Windows の場合

```
%PDDIR%\plugin\_phsgml\conf\pdplgsys
```

ユニット単位

ユニット単位に定義情報を設定するユーザは、プラグインユニット定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phsgml/conf/pdplgutsys
```

- Windows の場合

```
%PDDIR%\plugin\_phsgml\conf\pdplgutsys
```

サーバ単位 (BES または SDS 単位)

各サーバ単位に定義情報を設定するユーザは、プラグインサーバ環境情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- HiRDB/ パラレルサーバの場合

UNIX の場合

```
$PDDIR/plugin/_phsgml/conf/BES 名称
```

Windows の場合

```
%PDDIR%\plugin\_phsgml\conf\BES 名称
```

- HiRDB/ シングルサーバの場合

UNIX の場合

```
$PDDIR/plugin/_phsgml/conf/SDS 名称
```

Windows の場合

```
%PDDIR%\plugin¥_phsgml¥conf¥SDS 名称
```

優先順位は、サーバ単位 > ユニット単位 > システム単位です。

記述誤りなどが検出された場合、メッセージログに警告メッセージが出力され、誤りのある定義ファイルは無視されます。環境定義ファイルがないか、環境定義ファイルはあるが設定されていない定義がある場合は、エラーではなく、定義値なしとして処理されます。最終的に有効な値がない場合は、システムのデフォルトで動作します。

SGML プラグインの環境定義ファイルの定義例を次の図に示します。

図 2-35 SGML プラグインの環境定義ファイルの定義例

```
###SGML プラグイン環境定義ファイル
set phs_search_wait=nowait      # 登録・検索同時実行指定      #
set phs_txt_org=Y               # 原文書保持モード指定      #
set phs_document_size_max=xxxx # 登録文書サイズ上限値      #
set phs_errorcode_detail=off    # エラーコード詳細化指定    #
```

注意事項

- 各設定項目の行末は改行で区切ってください。
- 設定項目の途中を改行で区切らないでください。
- 1 行に設定できる文字列長は 1,023 バイトです。
- # (シャープ) 以降は、コメント文として扱われます。
- 各行は、空行なしで詰めて記述してください。

(2) 設定できる項目

SGML プラグインを動作させる環境情報を SGML プラグインの環境定義ファイルに定義することで、ユーザごとの環境を設定できます。設定できる定義項目を次の表に示します。なお、設定値項目には、ユーザごとに、必要に応じて取得できる最大値を設定します。

表 2-10 SGML プラグインの環境定義ファイルに設定できる定義項目

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を設定できる単位 |
|-------|-------------|---------------------|--|------------|
| 登録・検索 | 登録・検索同時実行指定 | set phs_search_wait | wait, nowait, または nowait_with_update デフォルト: wait | サーバ |

2. Text Search Plug-in システムの構築

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を設定できる単位 |
|-------|---------------|---------------------------|--|-----------------|
| 登録 | 原文書データ保持モード指定 | set phs_txt_org | Y または N デフォルト: Y | システム, ユニット, サーバ |
| | 登録文書サイズ上限値指定 | set phs_document_size_max | 最大値: 5,120 キロバイト 最小値: 1 キロバイト デフォルト: 5,120 キロバイト | システム, ユニット, サーバ |
| エラー処理 | エラーコード詳細化指定 | set phs_errorcode_detail | off または on デフォルト: off | システム, ユニット, サーバ |

登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行指定

登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行を使用するかどうかを指定します。

なお, この項目を設定できる単位は, サーバ単位だけです。

形式

```
set phs_search_wait=wait | nowait | nowait_with_update
```

- wait
同時実行を使用しない。
- nowait
登録・検索同時実行を使用する。
- nowait_with_update
登録・検索同時実行機能に加えて, 分割遅延登録・更新系 SQL 同時実行を使用する。

原文書データ保持モード指定

登録テキストの原文をデータベースに保持するかどうかを指定します。原文書を保持しない場合, 抽象データ型関数 `extracts` を発行するときに, 原文書出力機能を使用できなくなりますが, RD エリア使用量を節約できます。

形式

```
set phs_txt_org=Y | N
```

- Y
原文書を保持します。
- N
原文書を保持しません。

登録文書サイズ上限値指定

データベースに登録できる文書サイズの上限值を指定します。

形式

```
set phs_document_size_max=xxxx
```

エラーコード詳細化指定

SQL 連絡領域にメッセージが返却される場合に、エラーコードを詳細モードで出力するかどうかを指定します。

詳細モードで出力しない場合、SQLCODE=-997 のエラーとなります。詳細モードで出力する場合、Text Search Plugin のエラーコードが出力されます。エラーコードの詳細は「付録 F.1 SQL 連絡領域に出力されるメッセージ」を参照してください。

形式

```
set phs_errorcode_detail=off | on
```

- off
詳細化モードで出力しません。
- on
詳細化モードで出力します。

(3) 有効範囲

環境定義ファイルに定義した環境情報は、サーバ開始から終了まで有効です。HiRDB 稼働中に環境情報を変更した場合は、正しく動作しない場合がありますので、変更しないでください。

2.6.2 n-gram インデクスプラグインの環境設定

(1) 環境情報の定義方法

n-gram インデクスプラグインの環境定義ファイルは、サンプルファイルとして提供されています。なお、HiRDB/ シングルサーバの場合は、次のどの単位で設定しても同じ意味になります。

n-gram インデクスプラグインの環境情報を設定できる単位を次に示します。

システム単位

システム単位に定義情報を設定するユーザは、プラグインシステム定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phngram/conf/pdplgsys
```

- Windows の場合

```
%PDDIR%\plugin¥_phngram¥conf¥pdplgsys
```

ユニット単位

ユニット単位に定義情報を設定するユーザは、プラグインユニット定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

2. Text Search Plug-in システムの構築

- UNIX の場合

```
$PDDIR/plugin/_phngram/conf/pdplgutsys
```

- Windows の場合

```
%PDDIR%\plugin\_phngram\conf\pdplgutsys
```

サーバ単位

各サーバ単位に定義情報を設定するユーザは、プラグインサーバ環境情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- HiRDB/ パラレルサーバの場合

UNIX の場合

```
$PDDIR/plugin/_phngram/conf/BES 名称
```

Windows の場合

```
%PDDIR%\plugin\_phngram\conf\BES 名称
```

- HiRDB/ シングルサーバの場合

UNIX の場合

```
$PDDIR/plugin/_phngram/conf/SDS 名称
```

Windows の場合

```
%PDDIR%\plugin\_phngram\conf\SDS 名称
```

優先順位は、サーバ単位 > ユニット単位 > システム単位です。

記述誤りなどが検出された場合、メッセージログに警告メッセージを出力し、誤りのある定義ファイルは無視されます。環境定義ファイルがないか、環境定義ファイルはあるが設定されていない定義がある場合は、エラーではなく、定義値なしとして処理します。最終的に有効な値がない場合は、システムのデフォルトで動作します。

環境定義ファイルの定義例を次の図に示します。

図 2-36 n-gram インデクスプラグインの環境定義ファイルの定義例

```

###n-gramインデクスプラグイン環境定義ファイル
set phn_mem_build=xxxx # 一括登録時の最大メモリサイズ #
set phn_mem_insert=xxxx # 1件登録時の最大メモリサイズ #
set phn_mem_search=xxxx # 検索時最大ワークメモリサイズ #
set phn_dir_search=xxxxxxxxxx # 検索用ワークファイルディレクトリ #
set phn_opt_search=memory # 検索オプション #
set phn_mem_expand=xxxx # 同義語異表記展開最大メモリサイズ #
set phn_increment_practical_use=off # 異表記展開検索時の部分展開検索モード #
set phn_opt_score=boolean # スコア算出法指定 #
set phn_index_search=xxx # 検索抑止インデクスサイズ #
set phn_1gram_proximity_search= on # 1文字近傍検索実行可否 #
set phn_concept_terms=xxx # 概念検索用 検索用特徴ターム数指定 #
set phn_concept_idf=xxx # 概念検索用 特徴ターム抽出しきい値 #
set phn_seed_bound_for_concept_rule=xxx # 概念検索用 有効種文章サイズ #
set phn_lock_pool_size=xxxx # 1RDIリ当り排他制御用共用メモリサイズ #
set phn_delay_over=continue # 分割遅延登録用ファイルオーバ #
set phn_global_buffer_mode=0 # インデクス用グローバルバッファ最適化モード指定 #

```

注意事項

- 各設定項目の行末は改行で区切ってください。
- 1行に1項目だけ指定できます。2項目以上指定した場合は、2項目目以降は無視されます。
- 設定項目の途中を改行で区切らないでください。
- 検索用ワークファイルを使用して実行した検索が異常終了した場合、検索用ワークファイルディレクトリに不要なファイルが残ることがあります。これらのファイルは定期的にユーザが削除してください。
- 1行に設定できる文字列長は1,023バイトです。
- # (シャープ)以降は、コメント文として扱われます。
- 各行は、空行なしで詰めて記述してください。

(2) 設定できる項目

n-gram インデクスプラグインを動作させる環境情報を n-gram インデクスプラグインの環境定義ファイルに定義することで、ユーザごとの環境を設定できます。設定できる定義項目を次の表に示します。なお、各設定値項目には、ユーザごとに、必要に応じて取得できる最大値を設定します。

表 2-11 n-gram インデクスプラグインの環境定義ファイルに設定できる定義項目

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を設定できる単位 |
|----|-----------------------|-----------------------|---|---------------------|
| 登録 | 一括登録時 最大メモリ サイズ | set phn_mem_build | 最大値：2,097,152 キロバイト 最小値：30,720 キロバイト デフォルト：51,200 キロバイト | システム, ユ ニット, サーバ |
| | 1件登録時 最大メモリ サイズ | set phn_mem_insert | 最大値：2,097,152 キロバイト 最小値：3,072 キロバイト デフォルト：51,200 キロバイト | システム, ユ ニット, サーバ |

2. Text Search Plug-in システムの構築

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を設定できる単位 |
|-----------------|--------------------------|--|--|-----------------|
| 全文検索 および概念検索 | 検索時最大ワークメモリサイズ | set phn_mem_search | 最大値：2,097,152 キロバイト 最小値：2,048 キロバイト デフォルト：20,480 キロバイト | システム, ユニット, サーバ |
| | 検索用ワークファイルディレクトリ | set phn_dir_search | 最大長：200 バイト (¥0 を含む) 最小長：1 バイト UNIX の場合のデフォルト： HiRDB 運用ディレクトリ/ plugin/_phngram/tmp Windows の場合のデフォルト： HiRDB 運用フォルダ ¥plugin¥_phngram¥tmp | システム, ユニット, サーバ |
| | 検索オプション | set phn_opt_search | memory または file デフォルト：memory | システム, ユニット, サーバ |
| | 同義語異表記展開用最大メモリサイズ | set phn_mem_expand | 最大値：2,097,152 キロバイト 最小値：1 キロバイト デフォルト：0 (無制限) | システム, ユニット, サーバ |
| | 異表記展開検索時の部分展開検索モード | set phn_increment_practical_use | on または off デフォルト：off | システム, ユニット, サーバ |
| | スコア算出法指定 | set phn_opt_score | boolean または total デフォルト：boolean | システム, ユニット, サーバ |
| | 検索抑止インデックスサイズ | set phn_index_search | 最大値：2,097,152 キロバイト 最小値：64 キロバイト デフォルト：2,097,152 キロバイト | システム, ユニット, サーバ |
| 全文検索 | 1文字近傍検索実行可否 | set phn_lgram_proximity_search | on または off デフォルト：on | システム, ユニット, サーバ |
| 概念検索 | 検索用特徴ターム数 | set phn_concept_terms | 最大値：30 個 最小値：1 個 デフォルト：10 個 | システム, ユニット, サーバ |
| | 特徴ターム抽出しきい値 | set phn_concept_idf | 最大値：100% 最小値：1% デフォルト：50% | システム, ユニット, サーバ |
| | 有効種文章サイズ | set phn_seed_bound_for_concept_rule | 最大値：5,120 キロバイト 最小値：0 キロバイト デフォルト：5,120 キロバイト | システム, ユニット, サーバ |
| 登録・検索同時実行 | 1RD エリア当たりの排他制御用共用メモリサイズ | set phn_lock_pool_size | 最大値：1,024 キロバイト 最小値：8 キロバイト デフォルト：8 キロバイト | サーバ |

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を設定できる単位 |
|--------|-------------------------|-------------------------------|---------------------------------------|-----------------|
| 分割遅延登録 | 分割遅延登録用ファイルオーバー時の処理 | set phn_delay_over | continue または error デフォルト: continue | システム, ユニット, サーバ |
| 運用 | インデクス用グローバルバッファ最適化モード指定 | set phn_global_buffer_mode | 0 または 1 デフォルト: 0 | システム, ユニット, サーバ |

一括登録時最大メモリサイズ

n-gram インデクスに対する一括追加 (CREATE INDEX, pdload) をするとき使用する最大メモリサイズ (キロバイト) です。

形式

```
set phn_mem_build=xxxx
```

1件登録時最大メモリサイズ

n-gram インデクスに 1 件追加 (INSERT 文, UPDATE 文, pdload の SYNC モード) をするとき使用する最大メモリサイズ (キロバイト) です。

形式

```
set phn_mem_insert=xxxx
```

検索時最大ワークメモリサイズ

検索処理をするときに使用する作業用メモリの最大メモリサイズ (キロバイト) です。

形式

```
set phn_mem_search=xxxx
```

検索用ワークファイルディレクトリ

検索オプションに, "set phn_opt_search=file" を指定した場合, 検索に使用するワークファイルを作成するディレクトリ名を絶対パスで指定します。Text Search Plug-in は, 検索開始時にファイル名が "swk" で始まるユニークなファイルを指定されたディレクトリに作成し, 検索終了時に削除します。

検索が複数同時実行された場合, ディレクトリ競合による性能低下のおそれがあります。そのため, 同一の BES に対して複数のユーザが同時に検索する運用の場合は, ファイルではなくメモリの使用をお勧めします。検索処理が異常終了した場合, 作成されたワークファイルは削除されないので, HiRDB 管理者が削除してください。

形式

```
set phn_dir_search=xxxxxxxxxxxxxxxx
```

検索オプション

検索時の動作方法を指定します。

2. Text Search Plug-in システムの構築

形式

```
set phn_opt_search=memory | file
```

- memory
メモリ上で動作します。set phn_mem_search で指定したメモリサイズを超える場合は、メモリ不足エラーとなります。
- file
メモリおよび検索用ワークファイルで動作します。

同義語異表記展開用最大メモリサイズ

同義語異表記展開処理をするときに使用する最大メモリサイズ(キロバイト)です。0を指定した場合は、無制限となります。

同義語異表記展開用メモリサイズの見積もりについては、「2.3 リソースの見積もり」を参照してください。

なお、本パラメータは異表記展開検索時の部分展開検索モードが"on"(全展開検索モード)の場合に有効となります。

形式

```
set phn_mem_expand=xxxx
```

異表記展開検索時の部分展開検索モード

異表記展開検索時に、展開処理を部分的に実行するか、全面展開するかを指定します。この指定は、環境がシステム、ユニット、またはサーバの場合に有効です。

形式

```
set phn_increment_practical_use = on | off
```

- on
全展開検索モードで検索します。
- off
部分展開検索モードで検索します。

かたかな異表記展開検索を実行する場合は、部分展開検索モード指定(off指定)を推奨します。

スコア算出法指定

検索時のスコア算出法を指定します。

形式

```
set phn_opt_score=boolean | total
```

- boolean
論理演算時に、検索条件が論理積の場合は演算対象内で最小のスコアを最終スコアとし、論理和の場合は演算対象内で最大のスコアを最終スコアとします。
- total
論理演算時に、検索条件が論理積、論理和のどちらの場合でも、演算対象となるスコアの総和の値を最終スコアとします。

検索抑止インデクスサイズ

検索で使用するインデクスサイズの上限值を指定します。検索時に指定した上限値を超えるインデクスを参照すると、SQL 連絡領域に次のメッセージが出力されます。

```
Index data for search exceeds maximum size.(XX)
```

メッセージの詳細については、「付録 F エラーメッセージ」を参照してください。

形式

```
set phn_index_search = xxx
```

1 文字近傍検索実行可否

1 文字のインデクスを使用する近傍検索や、「A*A」のような 1 文字のインデクスを使用する前方後方一致のワイルドカード検索の実行を可能にするかを指定します。なお、これらの検索は、参照するインデクスが多くなり、システムに負荷を与えることがあります。このため、これらの検索を実行する必要がない場合は、「off」を指定してください。

形式

```
set phn_lgram_proximity_search = on | off
```

- on

1 文字のインデクスを使用する近傍検索や前方後方一致のワイルドカード検索を実行できます。

- off

1 文字のインデクスを使用する近傍検索や前方後方一致のワイルドカード検索を実行するとエラーリターンして、SQL 連絡領域に次のメッセージが出力されます。

```
Cannot specify term of one character in proximity.
```

メッセージの詳細については、「付録 F エラーメッセージ」を参照してください。

検索用特徴ターム数

概念検索で、検索条件の種文章から抽出した特徴タームのうち、検索で使用するターム（検索用特徴ターム）の個数の上限値を指定します。ターム数を増減することで、検索の幅を調整します。

より多くのタームを使用して幅広く検索したい場合は、検索用特徴ターム数の指定値を大きくします。ただし、ターム数を増やすと、ヒット件数が増えたり、検索に時間が掛かったりする場合があります。

形式

```
set phn_concept_terms=XXX
```

特徴ターム抽出しきい値

2. Text Search Plug-in システムの構築

概念検索で、検索条件の種文章から検索用特徴タームを抽出する場合に、データベース中（全文書中）で出現頻度が高い（種文章を特徴付ける度合いが低い）特徴タームを除くためのしきい値です。抽出しきい値には出現頻度が何%以下の特徴タームを検索用特徴タームの候補にするかを指定します。

出現頻度が高いタームを除いてヒット件数をできるだけ抑えたい場合は、特徴ターム抽出しきい値の指定値を小さくします。ただし、抽出しきい値を小さくすると、種文章の内容によっては抽出されたタームがすべて除外対象となり、ヒット件数が0件となる場合があります。

例えば、抽出しきい値を10%に設定した場合、全体の10%を超える文書に出現する特徴タームは、検索用特徴タームの候補から除外されます。

形式

```
set phn_concept_idf=XXX
```

有効種文章サイズ

概念検索を有効にする種文章のサイズ（キロバイト）を指定します。

Text Search Plug-in Version 2以降、Version 1よりも幅広く検索用特徴タームを抽出し、データベース内を検索します。種文章のサイズがこの項目に指定した値を超えた場合は、Version 1のバージョンの概念検索を実行します。無条件でVersion 1のバージョンの概念検索を実行したい場合は、0を指定してください。

形式

```
set phn_seed_bound_for_concept_rule=xxx
```

1RD エリア当たりの排他制御用共用メモリサイズ

登録・検索同時実行時の排他制御に使用する共用メモリのサイズ（キロバイト）を指定します。サイズは8キロバイトの倍数で指定してください。8キロバイトの倍数でない値が指定された場合は、8キロバイト単位で切り下げられます。ただし、8キロバイトより小さい値が指定された場合は、8キロバイト（デフォルト）になります。

なお、この項目を設定できる単位は、サーバ単位だけです。

形式

```
set phn_lock_pool_size=xxxx
```

分割遅延登録用ファイルオーバー時の処理

インデクス分割遅延登録時に、指定された分割登録用の領域を使い切った場合の処理を指定します。

形式

```
set phn_delay_over=continue | error
```

- continue

登録要求された文書は遅延せず、即時にインデクスを作成して登録します。

- error

エラーリターンします。

インデクス用グローバルバッファ最適化モード指定

インデクスにアクセスするときの HiRDB BLOB 用グローバルバッファを最適化するかどうか指定します。最適化した場合、検索性能の向上が期待できます。

形式

```
set phn_global_buffer_mode=0 | 1
```

- 0

最適化しません。

- 1

最適化します。

検索用の管理データをファイル単位で優先してバッファリングします。各ファイルの優先付けについては、「5.2.1 グローバルバッファ管理の運用」を参照してください。

(3) 有効範囲

環境定義ファイルに定義した環境情報は、サーバ開始から終了まで有効です。HiRDB稼働中に環境情報を変更した場合は、正しく動作しない場合がありますので変更しないでください。

2.7 文字列検索プラグインのデータベースの作成

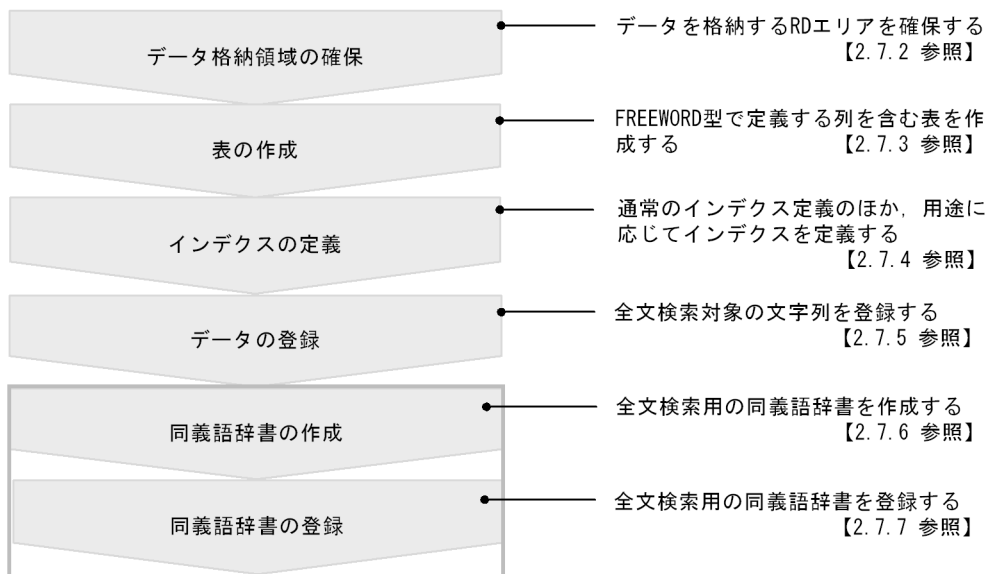
この節では、文字列検索プラグインのデータベースの作成について説明します。

なお、Text Search Plug-in で必要となるデータベースを作成するには、HiRDB が提供しているユティリティおよび運用コマンドを使用する場合があります。これらのユティリティおよび運用コマンドの使用方法については、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。また、Text Search Plug-in のユティリティの使用方法については、「3. Text Search Plug-in のユティリティ」を参照してください。

2.7.1 文字列検索プラグインのデータベース作成の流れ

Text Search Plug-in の導入後、表作成からデータベースへの登録までの操作方法の流れを次の図に示します。

図 2-37 データベース作成の流れ（文字列検索プラグイン）



の内容は、表の定義から同義語展開検索機能を使用するまでに設定してください。

以降、上記の流れに従って、データベースの作成方法について説明します。

2.7.2 データ格納領域の確保

FREEWORD 型を含む表データの格納領域、および IXFREEWORD インデクス情報の格納領域を確保します。データ格納領域の確保には、HiRDB の「pdinit (データベース初期設定ユティリティ)」または「pdmod (データベース構成変更ユティリティ)」の create rdarea 文を使用します。Text Search Plug-in では、FREEWORD 型を含む表の基本データ格納領域および IXFREEWORD インデクス情報格納領域の確保が必要です。

FREEWORD 型を含む表の基本データ格納領域にはユーザ用 RD エリアを、IXFREEWORD インデクス情報格納領域にはユーザ LOB 用 RD エリアを確保してください。

create rdarea 文の指定例を図 2-38 および図 2-39 に示します。

図 2-38 create rdarea 文の例 (UNIX の場合)

```

/* テーブル基本データ格納用RDエリア */
create rdarea PDUSER04 for user used by public      /* RDエリア名をPDUSER04に設定 */
server name bes1                                   /* サーバ名の指定bes1      */
storage control segment 1 pages
file name "/DB/DB01/LOB14"
initial 5000 segments ;

/* IXFREEWORDインデクス情報格納用RDエリア */
create rdarea PDUSER05 for LOB used by public      /* RDエリア名をPDUSER05に設定 */
server name bes1                                   /* サーバ名の指定bes1      */
storage control segment 1 pages
file name "/DB/DB01/LOB15"
initial 5000 segments ;

```

図 2-39 create rdarea 文の例 (Windows の場合)

```

/* テーブル基本データ格納用RDエリア */
create rdarea PDUSER04 for user used by public      /* RDエリア名をPDUSER04に設定 */
server name bes1                                   /* サーバ名の指定bes1      */
storage control segment 1 pages
file name "c:¥DB¥DB01¥LOB14"
initial 5000 segments ;

/* FREEWORDインデクス情報格納用RDエリア */
create rdarea PDUSER05 for LOB used by public      /* RDエリア名をPDUSER05に設定 */
server name bes1                                   /* サーバ名の指定bes1      */
storage control segment 1 pages
file name "c:¥DB¥DB01¥LOB15"
initial 5000 segments ;

```

2.7.3 表の作成

表の作成には、定義系 SQL の CREATE TABLE を使用します。CREATE TABLE では、文字列データを登録する列に対する列定義のデータ型として、FREEWORD と指定して

ください。なお、CREATE TABLE を実行するときは、データを格納する RD エリアの確保が必要です。

CREATE TABLE の指定例を次の図に示します。

図 2-40 表の定義例

```
CREATE TABLE reports
(
  title CHAR(32),      ←title列の定義
  news DATE,          ←news列の定義
  author CHAR(32),    ←author列の定義
  fword FREEWORD      ←fword(FREEWORD型)列の定義
) IN (PDUUSER04);    ←表格納RDエリアの指定
```

2.7.4 インデクスの定義

インデクスデータを作成してデータを登録します。データ登録後に CREATE INDEX を実行すると、登録済みの文字列データから IXFREEWORD インデクスが一括作成されず。

文字列検索プラグインでのインデクスの作成には、通常のインデクス定義、および文書の登録・更新時のインデクス作成方法の定義があります。各定義方法について次に説明します。

(1) 通常のインデクス定義

全文検索用のインデクス作成情報を定義するには、定義系 SQL の CREATE INDEX を使用します。この定義をすることで、全文検索機能が使用できるようになります。

CREATE INDEX のインデクス型識別子には IXFREEWORD と指定してください。また、RD エリア名には、「2.7.2 データ格納領域の確保」で IXFREEWORD インデクス情報の格納領域として確保した RD エリアを指定してください。IXFREEWORD プラグインで使用するインデクスの RD エリアは、ログ取得モードとなります。CREATE INDEX の詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

なお、CREATE INDEX 実行時に指定するプラグインオプションには複数の定義が指定できます。

(例)

複数の定義を指定する場合は、次のように定義ごとに、(半角コンマ)で区切って指定します。

```
PLUGIN '定義,定義,定義,...'
```

インデクスの定義例を次の図に示します。

図 2-41 全文検索機能使用時のインデクス定義例

| | |
|---------------------------|------------------|
| CREATE INDEX REPORTSindex | ← インデクス識別子の指定 |
| USING TYPE IXFREEWORD | ← インデクス型識別子の指定 |
| ON reports (fword) | ← インデクスを定義する列の指定 |
| IN (PDUER05); | ← 格納RDエリアの指定 |

(2) インデクスファイル自動拡張機能のインデクス定義

インデクスファイル自動拡張機能を利用すると、インデクスファイル作成時の空き容量を、各論理ファイルの拡張領域として使用できるようになります。これによって RD エリアを有効利用できますので、インデクスファイル自動拡張機能の使用を推奨します。この機能を使用する場合は、プラグインオプションに次の定義が必要です。

```
PLUGIN 'FLEX_AREA'
```

この機能を使用すると自動的にインデクスが拡張するため、インデクスの拡張ユティリティ (phxexpnd) を実行する必要はありません。この機能を利用した場合にインデクスの拡張ユティリティを実行したときは、無視されます。

インデクスファイル自動拡張機能使用時のインデクス定義例を、次の図に示します。

図 2-42 インデクスファイル自動拡張機能使用時のインデクス定義例

| | |
|---------------------------|------------------------|
| CREATE INDEX REPORTSindex | ← インデクス識別子の指定 |
| USING TYPE IXFREEWORD | ← インデクス型識別子の指定 |
| ON reports (fword) | ← インデクスを定義する列の指定 |
| IN (PDUER05) | ← 格納RDエリアの指定 |
| PLUGIN 'FLEX_AREA' ; | ← インデクスファイル自動拡張機能使用の指定 |

(3) 大小文字・全角文字半角文字統一機能のインデクス定義

大文字と小文字、または全角文字と半角文字の違いを区別しないで検索を実行したい場合には、大小文字・全角文字半角文字統一機能を使用します。

(a) 大小文字・全角文字半角文字統一機能の定義方法

大小文字・全角文字半角文字統一機能を使用する場合には、CREATE INDEX のプラグインオプションに対して次の表に示す内容を定義します。

表 2-12 定義内容

| 定義内容 | 説明 |
|-------------|--|
| SAMECASE=ON | 英大文字と英小文字を区別しないでインデクスを作成します。 |
| SAMEWIDE=ON | 全角文字と半角文字の英数字、かたかな、記号を区別しないでインデクスを作成します。 |

2. Text Search Plug-in システムの構築

注

半角の英字で記述します。SAMECASE と SAMEWIDE を組み合わせて指定できます。

大小文字・全角文字半角文字統一機能使用時のインデクス定義例を図 2-43, 図 2-44, 図 2-45 に示します。

図 2-43 大小文字統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUER05)
  PLUGIN 'SAMECASE=ON';    ← 大小文字統一機能使用の指定
```

図 2-44 全角文字半角文字統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUER05)
  PLUGIN 'SAMEWIDE=ON';    ← 全角文字半角文字統一機能使用の指定
```

図 2-45 大小文字統一機能, および全角文字半角文字統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUER05)
  PLUGIN 'SAMECASE=ON, SAMEWIDE=ON';    ← 大小文字統一機能と全角文字半角文字統一機能を同時に指定
```

(b) 大小文字・全角文字半角文字統一機能使用時の注意

- 同義語展開指定をする場合, あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか, または同義語辞書移行ユーティリティ (「phcsyndicconv」) を実行して, 既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合, 同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。
- 全角文字半角文字統一機能を指定しないで, 近傍条件検索やワイルドカードに "?" を指定した検索を実行した場合, 半角文字のガ行, ザ行, ダ行, パ行, バ行, およびヴを 2 文字として扱います。しかし, 全角文字半角文字統一機能を指定した場合は 1 文字として扱うため注意が必要です。

(4) 拗音統一機能のインデクス定義

検索時に「や」と「ゃ」, 「よ」と「ょ」などの拗音を区別しないようにするには, 拗音

統一機能を使用します。

(a) 拗音統一機能の定義方法

拗音統一機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義を追加してください。

```
PLUGIN 'SAMEY=ON'
```

拗音統一機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-46 拗音統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUSER05)
  PLUGIN 'SAMEY=ON';    ← 拗音統一機能使用の指定
```

「SAMEY=ON」を指定すると、次の拗音などが統一されます。

●全角文字

あ, い, う, え, お, つ, や, ゆ, よ, ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

●半角文字

ア, イ, ウ, エ, オ, ツ, ヤ, ユ, ヨ

(b) 拗音統一機能使用時の注意

- 同義語展開指定をする場合、あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか、または同義語辞書移行ユティリティ（「phcsyndicconv」）を実行して、既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合、同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。

(5) 濁音統一機能のインデクス定義

検索時に「バ」と「ハ[゜]」、「ば」や「は[゜]」のように濁音や半濁音を区別しないようにするには、濁音統一機能を使用します。

(a) 濁音統一機能の定義方法

濁音統一機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義を追加してください。

```
PLUGIN 'SAMED=ON'
```

濁音統一機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-47 濁音統一機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUSER05)
  PLUGIN 'SAMED=ON';    ← 濁音統一機能使用の指定
```

「SAMED=ON」を指定すると、濁音、半濁音が統一されます。

(b) 濁音統一機能使用時の注意

- 同義語展開指定をする場合、あらかじめ Text Search Plug-in 02-01 以降のバージョンで同義語辞書を作成するか、または同義語辞書移行ユティリティ (「phcsyndicconv」) を実行して、既存の同義語辞書を移行する必要があります。
- 02-01 より前のバージョンで作成した同義語辞書を使用した場合、同義語異表記展開を指定しても同義語辞書に登録した同義語で検索できない場合があります。
- 全角文字半角文字統一機能を指定しないで、近傍条件検索やワイルドカードに "? " を指定した検索を実行した場合、半角文字の濁音で表示できる文字を 2 文字として扱います。しかし、全角文字半角文字統一機能を指定した場合は 1 文字として扱うため注意が必要です。

(6) 削除文字指定機能のインデクス定義

検索に使用しない文字を除いて検索性を向上させるには、削除文字指定機能を使用します。

(a) 削除文字指定機能の定義方法

削除文字指定機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義を追加してください。

```
PLUGIN 'DELcode=ON'
```

削除文字指定機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-48 削除文字指定機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(fword)
  IN (PDUSER05)
  PLUGIN 'DELcode=ON';    ← 削除文字指定機能使用の指定
```

「DELcode=ON」を指定すると、次の文字が削除されます。

全角文字の "。(句点), "。(読点), 全角スペース,

半角文字の "。(句点), "(読点), 半角スペース (0x20),
改行コード (0x0a, 0x0d), タブコード (0x09)

(b) 削除文字指定機能使用時の注意

- 近傍検索で指定された語句リストの間に削除文字が含まれる場合、距離の指定方法によってはヒットしないことがあります。
例えば、登録データが「COMPUTER NETWORK」の場合、「COMPUTER」と「NETWORK」の間の距離は0となります。
- 削除文字を指定したインデクスに対して検索する場合、登録データから削除文字が取り除かれていることに注意してください。
- 検索タームおよびすべての異表記展開・同義語異表記展開結果が、削除文字だけで構成されている場合、0件ヒットとなります。

(7) 文字単位インデクス作成機能使用時のインデクス定義

インデクス容量を削減する方法として、1文字ずつインデクスを作成することができます。この機能を文字単位インデクス(1gram インデクス)作成機能といいます。

(a) 文字単位インデクス作成機能の定義方法

文字単位インデクス作成機能を使用する場合は、CREATE INDEX のプラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'INDEX_GRAM1'
```

文字単位インデクス作成機能を使用する場合の、インデクスの定義例を次の図に示します。

図 2-49 文字単位インデクス作成機能使用時のインデクス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports (doc)
  IN (PDUER03)
  PLUGIN 'INDEX_GRAM1'; ← 文字単位インデクス作成機能使用の指定
```

(b) 文字単位インデクス作成機能使用時の注意事項

- 文字単位インデクス作成機能は、次の機能と同時に指定することはできません。
 - 概念検索用インデクスの作成
 - インデクス作成ルール追加機能の使用
- 文字単位インデクスだけ作成した場合、インデクス容量が削減される反面、登録データの増加に伴い、検索性能が低下します。このため、大量データを登録する運用にこの機能を適用する場合は注意が必要です。

(8) インデクス作成方法指定

表の FREEWORD 型列に文字列データを登録している状態で CREATE INDEX を実行する場合、CREATE INDEX 実行時にインデクス作成とインデクスへの文字列データ登録を実行する方法（即時登録）か、CREATE INDEX 実行時にはインデクスに反映しないで文書やデータの登録だけを実行し、「phxstartidx（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の文字列データ登録を実行する方法（初期分割登録または分割遅延登録）のどちらかのインデクス作成方法を指定できます。運用形態に合わせて指定してください。

(a) インデクス作成方法指定の定義方法

インデクス作成方法指定を使用する場合、CREATE INDEX のプラグインオプションに即時登録、初期分割登録、または分割遅延登録の定義を追加します。インデクス作成方法指定の定義を省略した場合、即時登録が適用されます。

即時登録

表に登録されている文字列データについて、CREATE INDEX 実行時にインデクス作成とインデクスへの文字列データ登録を実行します。

即時登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加、または省略します。

```
PLUGIN 'CREATEMODE=0'
```

初期分割登録

表に登録されている文字列データについて、CREATE INDEX 実行時にはインデクスへ反映しません。「phxstartidx（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の文字列データ登録を実行します。

初期分割登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=1'
```

初期分割登録使用時のインデクスの定義例を次の図に示します。

図 2-50 初期分割登録使用時のインデクス定義例

```
CREATE INDEX REPORTindex
USING TYPE IXFREEWORD
ON reports (fword)
IN (PDUSER05)
PLUGIN 'CREATEMODE=1'; ← 初期分割登録の指定
```

分割遅延登録

表に登録されている文書について、CREATE INDEX 実行時にはインデクスへ反映しません。「phxstartidx（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の文書登録を実行します。分割遅延登録を使用する場合、インデクス分割遅延登録用の領域が必要となります。

分割遅延登録を使用する場合、CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=2, DELAY_SIZE=インデックス分割遅延登録用領域サイズ (キロバイト)'
```

インデックス分割遅延登録用領域サイズの見積もりについては、「2.3.2(2)(b) ユーザ LOB 用 RD エリア容量」を参照してください。また、分割遅延登録を設定していないインデックスに対しては、「phxmodidx (インデックス情報変更ユティリティ)」を実行することで分割遅延登録を設定したインデックスに移行できます。指定できる最大値・最小値は次のとおりです。

- 最大値：129,584 キロバイト
- 最小値：2,608 キロバイト

分割遅延登録時のインデックスの定義例を次の図に示します。

図 2-51 分割遅延登録使用時のインデックス定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports (fword)
  IN (PDUSER05)
  PLUGIN 'CREATEMODE=2, DELAY_SIZE=10240'; ← 分割遅延登録の指定, インデックス分割遅延登録
  用領域サイズを10,240(キロバイト)に設定
```

分割遅延登録への移行時に実行する、「phxmodidx (インデックス情報変更ユティリティ)」の指定例を次の図に示します。

図 2-52 分割遅延登録への移行時に実行する、インデックス情報変更ユティリティの指定例

```
phxmodidx -d REPORTindex -v "DELAY_SIZE=10240"; ← インデックス分割遅延登録用領域サイズ
  を10,240(キロバイト)に設定
```

分割遅延登録・更新系 SQL 同時実行

遅延登録実行時に、同時に「phxstartidx (インデックス作成開始ユティリティ)」を実行してインデックスへ文書を登録します。分割遅延登録・更新系 SQL 同時実行を利用する場合、遅延登録時に情報を退避しておく領域が必要です。CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN 'CREATEMODE=3, DELAY_SIZE=遅延登録時に情報を退避する領域サイズ (キロバイト)'
```

遅延登録時に情報を退避する領域サイズの見積もりについては、「2.3.2(2)(b) ユーザ LOB 用 RD エリア容量」を参照してください。

この退避領域に指定できる最大値・最小値は次のとおりです。

2. Text Search Plug-in システムの構築

- 最大値：129,584 キロバイト
- 最小値：12,832 キロバイト

分割遅延登録・更新系 SQL 同時実行のインデクスの定義例を次の図に示します。

図 2-53 分割遅延登録・更新系 SQL 同時実行のインデクスの定義例

```
CREATE INDEX REPORTindex
  USING TYPE IXFREEWORD
  ON reports(doc)
  IN (PDUUSER03)
  PLUGIN 'CREATEMODE=3, DELAY_SIZE=12832'; ← 分割遅延登録・インデクス作成同時実行の指定、
                                              遅延登録時に情報を退避させるスタック領域サ
                                              イズを12,832(キロバイト)に設定
```

(b) 初期分割登録・分割遅延登録使用時の注意

- CREATE INDEX 実行後、インデクスへ登録されていない文字列データのインデクスは、「phxstartidx (インデクス作成開始ユティリティ)」を実行して登録してください。インデクスへ登録されていない文字列データが存在するかどうかは、「phxidxls (インデクス情報の取得ユティリティ)」を実行して確認してください。
- インデクスへ登録されていない文字列データは、検索対象になりません。
- インデクスへ登録されていない文字列データが存在する場合、登録または更新した文字列データはインデクスへ登録されません。インデクスへ登録する場合は、「phxstartidx (インデクス作成開始ユティリティ)」を実行してください。
- インデクスが作成されていない文字列データが存在する場合、検索を実行しても、インデクスが作成されていない旨のメッセージは出力されません。
- 初期分割登録の場合、CREATE INDEX 実行後、インデクス作成開始ユティリティを実行して、すべての文字列データがインデクスへ登録された時点で、インデクス作成方法が初期分割登録から即時登録に切り替わります。以降、文字列データの登録または更新と同時にインデクスへの登録が行われるようになるため、文字列データの登録または更新時にインデクスへ登録する時間が掛かるようになります。
- 「phxmodidx (インデクス情報変更ユティリティ)」で移行したインデクスに対して初期化をした場合、変更した情報は無効になります。
- インデクス分割遅延登録用領域を使い切った場合は、遅延しているすべての文書をインデクスに反映するか、または「phxmodidx (インデクス情報変更ユティリティ)」を使用してインデクス分割遅延登録用領域を拡張してください。
- 分割遅延登録・更新系 SQL 同時実行をしない運用へ移行する場合、インデクスを削除してください。そのあと、運用に合わせてインデクスを再定義してください。
- 分割遅延登録・更新系 SQL 同時実行を行う運用にする場合、以下の機能が実行できません。
 - 初期分割登録 (CREATEMODE=1)
 - 時間指定でのインデクス作成
 - 遅延 / 即時の切り替え (phxdelay)
 - phxmodidx コマンドによる CREATEMODE=3 への移行

(9) UCS4 のインデクス定義

UCS2 の範囲を超える UCS4 の範囲の文字を使用するには、UCS4 用の定義が必要になります。

(a) インデクスの指定

UCS2 の範囲を超える UCS4 の範囲の文字を使用する場合は、プラグインオプションに次の定義の追加が必要です。

```
PLUGIN 'INDEX_UCS4'
```

(b) インデクス定義時の注意

- UCS4 用のインデクス定義の追加をしていない状態で UCS2 の範囲を超える文字を登録した場合は、該当文字は無視されます。
- UCS4 用のインデクス定義の追加をしていない状態で UCS2 の範囲を超える文字を検索した場合は、エラーとなります。
- HiRDB の文字コード種別が utf-8 以外の時に INDEX_UCS4 を指定した場合は、インデクス定義はエラーとなります。
- 以下のプラグインオプションは INDEX_UCS4 と同時に指定することができません。
 - 差分インデクス (SUB_INDEX)
 - 英文インデクス (ENGLISH, ENGLISH_STANDARD)

(c) UCS2 から UCS4 への移行

UCS2 と UCS4 はインデクスの構成が異なるため、UCS2 用のインデクスをそのまま UCS4 用のインデクスとして使用することはできません。

UCS2 用のインデクスを UCS4 用として使用する場合は、「phxmodidx (インデクス情報変更ユティリティ)」で変更項目に "INDEX_TYPE=UCS4" を指定して移行してください。

移行する前のインデクスに次の機能が指定されている場合は、移行コマンド実行時にエラーとなります。

- 英文インデクス (ENGLISH, ENGLISH_STANDARD)
- データベースの文字コード種別が utf-8 ではないデータベースの環境

(10) 英文検索機能のインデクス定義

英文に対する検索を行なう場合は、単語単位での検索を行なう英文インデクス定義をしてください。

英文インデクスの定義は次の方法があります。運用に応じてご使用ください。

- ENGLISH (単語一致)
- ENGLISH_STANDARD (単語一致, 前方一致)

図 2-54 英文検索機能を使用する場合のインデクス定義例（英語文字列検索）

| | |
|---|---|
| CREATE INDEX REPORTindex | ← インデクス識別子の指定 |
| USING TYPE IXFREEWORD | ← インデクス種別の指定 |
| ON reports(fword) | ← インデクスを定義する列の指定 |
| IN (PDUSER04) | ← 格納 R D エリアの指定 |
| PLUGIN 'ENGLISH_STANDARD, INDEXd1mt=INDEXd1mt.prm, NOindex=NOindex.prm, DELcode=DELcode.prm'; | ← 英文検索用インデクスを作成する指定, インデクスデリミタファイル名の指定, 不要語ファイル名の指定, 削除コードファイル名の指定 |

注意事項

英文インデクス定義は、次のインデクス定義と併用することはできません。

- SAMEWIDE=ON（全角文字半角文字統一機能）
- SAMEY=ON（拗音統一検索）
- INDEX_GRAM1（文字単位インデクス作成機能）

(a) 英文検索のインデクスに指定できる項目

英文検索機能で使用するインデクスの定義ファイルについて説明します。

インデクスデリミタファイル

機能

英文検索機能で使用するインデクスを作成する場合に必要な単語の区切り（デリミタ）を定義します。定義したインデクスデリミタに従って単語を抽出し、英文検索機能で使用するインデクスを作成します。抽出した単語が 64 文字（バイト）を超える場合は、65 文字（バイト）目以降を無視して、64 文字（バイト）目までの文字でインデクスを作成します。なお、インデクスデリミタは検索対象にできません。

指定形式

CREATE INDEX 文のプラグインオプションにインデクスデリミタファイルのファイル名を指定します。

INDEXd1mt=ファイル名

インデクスデリミタファイルは、CREATE INDEX を実行する前に、Text Search Plug-in をセットアップしたすべてのサーバマシンに対して、同一の内容で次に示すディレクトリに事前に作成しておく必要があります。

\$PDDIR/plugin/_phxfwr/idx_def

また、サンプルファイル（\$PDDIR/plugin/_phxfwr/sample/pdplgdlmt）を提供しています。

記述形式

- "インデクスデリミタ" [,"インデクスデリミタ"].....
- インデクスデリミタは,"(半角ダブルクォーテーション)で囲んで指定します。
 - インデクスデリミタに使用できる文字は,'0x21' ~ '0x7E'の範囲の1バイト文字コードです。
 - '0x20'(半角スペース),'0x09'(タブ)は,無条件にインデクスデリミタとなります。
 - 複数指定する場合は,"(半角ダブルクォーテーション)で囲んだインデクスデリミタを,(半角コンマ)で区切って指定します。
 - "(半角ダブルクォーテーション)をインデクスデリミタにする場合は,""と二つ続けて定義し,それを"(半角ダブルクォーテーション)で囲んで指定します。

注意事項

- インデクスデリミタファイルは省略できます。
- 省略した場合は,'0x20'(半角スペース),'0x09'(タブ),記号「'0x21' ~ '0x2F','0x3A' ~ '0x40','0x5B' ~ '0x60','0x7B' ~ '0x7E'」がインデクスデリミタとなります。
- 不要語,または削除コードに指定した文字コードをインデクスデリミタに指定しないでください。指定した場合は,インデクスデリミタ,削除コード,不要語,の優先順位で処理します。

不要語ファイル

機能

英文検索機能に使用するインデクスを作成する場合,検索に不要な単語を定義します。不要語については,検索用インデクスを作成しません。

指定形式

CREATE INDEX 文のプラグインオプションに不要語ファイルのファイル名を指定します。

NOindex=ファイル名

不要語ファイルは,CREATE INDEX を実行する前に,Text Search Plug-in をセットアップしたすべてのサーバマシンに対して,同一の内容で次のディレクトリに事前に作成しておく必要があります。

\$PDDIR/plugin/_phxfwr/idx_def

記述形式

["不要語" [,"不要語"].....]

- 不要語は,"(半角ダブルクォーテーション)で囲んで指定します。
- 不要語に使用できる文字は,'0x21' ~ '0x7E'の範囲1バイトの文字コードです。

2. Text Search Plug-in システムの構築

- 複数指定する場合は、" (半角ダブルクォーテーション) で囲んだ不要語を、(半角コンマ) で区切って指定します。不要語は 100 個まで指定できます。
- " (半角ダブルクォーテーション) を不要語にする場合は、"" と二つ続けて定義し、それを " (半角ダブルクォーテーション) で囲んで指定します。
- 不要語は 64 文字 (バイト) まで指定できます。64 文字 (バイト) を超えて指定した場合、65 文字 (バイト) 目以降を無視して、64 文字 (バイト) 目までの文字列を不要語とします。

注意事項

- 不要語ファイルは省略できます。省略した場合は、不要語なしでインデクスを作成します。
- インデクスデリミタ、または削除コードに指定した文字コードを不要語に指定しないようにしてください。指定した場合は、インデクスデリミタ、削除コード、不要語の優先順位で処理します。
- 検索条件で指定した検索タームに不要語を含む場合は、不要語を任意の 1 単語とみなして検索します。なお、検索条件に不要語だけを指定した場合は、入力条件エラーとなります。

削除コードファイル

機能

英文検索機能で使用するインデクスを作成する場合に、削除するコードを定義します。定義した削除コードはないものとして単語を抽出し、英文検索機能で使用するインデクスを作成します。

単語に - (ハイフン) のような不要なコードがある場合、削除してその前後を詰めて扱い、必要なコードだけでインデクスを作成できます。

指定形式

CREATE INDEX 文のプラグインオプションに削除コードファイルのファイル名を指定します。

DELcode=ファイル名

削除コードファイルは、CREATE INDEX を実行する前に、Text Search Plug-in をセットアップしたすべてのサーバマシンに対して、同じ内容で次のディレクトリに事前に作成しておく必要があります。

```
$PDDIR/plugin/_phxfwr/idx_def
```

記述形式

```
[" 削除コード " [, " 削除コード " ] . . . . . ]
```

- 削除コードは、" (半角ダブルクォーテーション) で囲んで指定します。
- 削除コードに使用できる文字は、'0x21' ~ '0x7E' の範囲の 1 バイト文字コードです。
- 複数指定する場合は、" (半角ダブルクォーテーション) で囲んだ削除コード

を ,(半角コンマ) で区切って指定します。

- "(半角ダブルクォーテーション) を削除コードにする場合は, "" と二つ続けて定義し, それを "(半角ダブルクォーテーション) で囲んで指定します。

注意事項

- 削除コードファイルは省略できます。省略した場合は, 削除コードなしでインデクスを作成します。
- インデクスデリミタ, または不要語に指定した文字コードを削除コードに指定しないようにしてください。指定した場合は, インデクスデリミタ, 削除コード, 不要語, の優先順位で処理します。

(11) 登録件数指定

本製品が推奨する文字列データサイズ (255 バイト) より小さい文書を大量に IXFREEWORD インデクス型プラグインに登録する場合, 登録文書数を管理する領域を拡張することがあります。その際には phidxls コマンドで表示されるインデクス情報の “ Available Work Area Size ” で示される領域を使用するため, この領域が不足すると RD エリア不足となり, RD エリアを追加する必要があります。本機能では, あらかじめユーザが最終目標とする文書数を管理できる領域を確保し, 前記の弊害をなくすことを目的としています。

登録件数指定を指定して IXFREEWORD インデクスを作成する場合には, CREATE INDEX のプラグインオプションに次の定義を追加します。

```
PLUGIN ' DOC_COUNT=登録件数 (件) '
```

登録件数に指定できる最大値・最小値は次のとおりです。

- 最大値: 16,000,000 件
- 最小値: 40,960 件

登録件数指定を省略した場合は, 次の (a)(b) の大きい方の値がデフォルト値となります。

(a) 40,960 件

(b) インデクス作成時の n-gram インデクス情報格納用 RD エリア容量 (バイト) / (100 × 32) 件

図 2-55 登録件数指定のインデクス定義例

```
CREATE INDEX REPORTindex  
  USING TYPE IXFREEWORD  
  ON reports(fword)  
  IN (PDUER05)  
  PLUGIN 'DOC_COUNT=40960';    ← 登録件数の指定
```

注意事項

- 表分割される場合は、各分割表ごとに指定された文書数が登録可能となります。
例えば、“DOC_COUNT=1000000”を表 A（分割表 A1/A2/A3 の 3 分割表）のインデクス作成時に指定した場合、各分割表ごとに 1,000,000 件のインデクスが作成できるので、表 A のインデクスとしては 3,000,000 件登録できます。
- UPDATE/DELETE によって無効となった文書も管理データとしては残ります。
例えば、10 件登録後に 5 件削除した場合、管理上は 10 件分の管理領域を必要とします。
また、10 件登録後に 5 件更新した場合、5 件削除後に 5 件追加する処理を行うため、15 件分の管理領域が必要となります。このような無効領域を解消するためには、インデクス再作成や表の再編成を行ってください。
- 登録件数が DOC_COUNT に指定した文書数を超えた場合、“Available Work Area Size” に空き容量がある限りは管理領域を拡張して登録を行います。

2.7.5 データの登録

(1) データの登録方法

作成した表にデータを登録するには、操作系 SQL の INSERT 文を使用します。登録する文字列データは、HiRDB をセットアップしたときの文字コード、および「付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字」に示す文字コードを使用してください。INSERT 文の詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

INSERT 文の指定例を次の図に示します。

図 2-56 INSERT 文の処理の例

```
EXEC SQL BEGIN DECLARE SECTION;
  struct {
    short      length;
    char       data[32000];
  } string;
  strcpy( string.data, char_ptr_pointing_to_a_fword_string);
  string.length = strlen( char_ptr_pointing_to_a_fword_string );
EXEC SQL END DECLARE SECTION;
EXEC SQL
  INSERT INTO reports(fword) VALUES( FREEWORD( :string AS VARCHAR(32000) ) );
```

また、HiRDB の「pdload (データベース作成ユーティリティ)」を使用すると、データの一括登録ができます。

(2) 登録・検索同時実行の設定

登録・検索同時実行を設定すると、データの登録と検索の同時実行ができます。同時実行を使用する利点を次に示します。

- ほかのユーザが更新中 (トランザクションが終了していない状態) の文書を検索できません。
- IXFREEWORD インデクス情報格納用 RD エリアに対してグローバルバッファを割り当てると、登録および検索性能が向上します。

ただし、同時実行できるのは、次の場合です。

- SQL による単数の更新系トランザクションと、複数の検索系トランザクションの同時実行
- インデクス未作成の場合、複数の更新系トランザクションの同時実行

(a) 指定方法

Text Search Plug-in の環境設定項目、および HiRDB のバックエンドサーバ定義またはシングルサーバ定義に、次の内容を設定します。Text Search Plug-in の環境設定項目については、「2.8 文字列検索プラグインのユーザ環境設定」を参照してください。また、HiRDB のバックエンドサーバ定義またはシングルサーバ定義の詳細については、マニュアル「HiRDB システム定義」を参照してください。

文字列検索プラグインの環境設定項目

- FREEWORD プラグインを使用している場合
FREEWORD プラグインの環境設定項目「登録・検索同時実行指定」で次のように指定します。

```
set phc_search_wait=nowait
```

- IXFREEWORD プラグインを使用している場合

2. Text Search Plug-in システムの構築

IXFREEWORD プラグインの環境設定項目「1RD エリア当たりの排他制御用共用メモリサイズ」で次のように指定します。

```
set phx_lock_pool_size=共用メモリサイズ
```

HiRDB のバックエンドサーバ定義またはシングルサーバ定義

- IXFREEWORD プラグインを使用している場合
IXFREEWORD プラグインで使用する共用メモリのサイズを指定します。
HiRDB のバックエンドサーバ定義またはシングルサーバ定義のオペランド「pdplgprm」は、次の形式で指定してください。

オペランドの指定形式

```
pdplgprm -n プラグイン名 -s 共用メモリサイズ
```

引数

-n プラグイン名

共用メモリを使用するプラグインの名称を指定します。ここでは、IXFREEWORD プラグインのプラグイン名「_phxfwr」を指定してください。

-s 共用メモリサイズ

IXFREEWORD プラグインで使用する共用メモリのサイズ（キロバイト）を指定します。

(b) 注意事項

- HiRDB が提供しているユティリティによる登録と検索系トランザクションは同時に実行できません。
- 「phxidxls（インデクス情報の取得ユティリティ）」および「phxstartidx（インデクス作成開始ユティリティ）」以外の Text Search Plug-in のユティリティと、検索系トランザクションは同時に実行できません。
- 同時実行を設定している場合、Text Search Plug-in のデータの整合性確保のために、一時的に排他制御をします。このため、同時実行を設定していない場合に比べて、排他制御によるオーバーヘッドが掛かります。
- "phc_search_wait=nowait" を設定した場合は、IXFREEWORD インデクスを作成していない場合でも共用メモリの定義をする必要があります。
- 同時実行する場合は、グローバルバッファを指定してください。実行性能が向上します。
- LOCK 文でテーブル全体を排他すると、ほかのトランザクションとの同時実行ができません。このため、分割遅延登録・更新系 SQL 同時実行を実施する場合は、LOCK 文でテーブル全体の排他を実施しないようにしてください。

(3) 分割遅延登録・更新系 SQL 同時実行の設定

分割遅延登録・更新系 SQL 同時実行を利用すると、(2) で説明した登録・検索同時実行に加えて、さらに、データ登録とインデクス作成を同時に実行できます。

(a) 指定方法

FREEWORD プラグインの環境設定項目「登録・検索同時実行」に、次のように指定します。

```
set phc_search_wait=nowait_with_update
```

(b) 注意事項

- LOCK 文でテーブル全体を排他すると、ほかのトランザクションとの同時実行ができません。このため、分割遅延登録・更新系 SQL 同時実行を実施する場合は、LOCK 文でテーブル全体の排他を実施しないようにしてください。
- 検索トランザクションの排他が更新系 SQL またはインデクス作成開始ユーティリティ (phxstartidx) と競合することを避けるため、検索 SQL 実行時は、SQL の排他オプションに WITHOUT LOCK NOWAIT を指定してください。

2.7.6 同義語辞書の作成

全文検索用の同義語辞書の作成方法は、文書検索プラグインの同義語辞書と同じです。文書検索プラグインの同義語辞書の作成方法については、「2.5.8 同義語辞書の作成」を参照してください。

2.7.7 同義語辞書の登録

同義語辞書の登録には、「phcsyndicreg (同義語辞書登録ユーティリティ)」を使用します。

同義語辞書登録ユーティリティの指定例を次に示します。

UNIX の場合

```
phcsyndicreg USERDIC01 /opt/def1/mydictionary.dic
```

Windows の場合

```
phcsyndicreg USERDIC01 c:¥data¥def1¥mydictionary.dic
```

注意事項

- 文書検索プラグインで登録した同義語辞書は、文字列検索プラグインでは使用できません。文書検索プラグインと同じ同義語辞書を使用する場合は、文書検索プラグインで登録した同義語辞書テキストファイルを、文字列検索プラグインにも同様に登録してください。

2.8 文字列検索プラグインのユーザ環境設定

この節では、Text Search Plug-in システムの文字列検索プラグインの環境をユーザごとに設定する方法について説明します。

2.8.1 FREEWORD プラグインの環境設定

(1) 環境情報の定義方法

FREEWORD プラグインの環境情報は、次の単位で設定できます。FREEWORD プラグインの環境定義ファイルは、サンプルファイルとして提供されています。

システム単位

システム単位に定義情報を設定するユーザは、プラグインシステム定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phcfwd/conf/pdplgsys
```

- Windows の場合

```
%PDDIR%\plugin\_phcfwd\conf\pdplgsys
```

ユニット単位

ユニット単位に定義情報を設定するユーザは、プラグインユニット定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phcfwd/conf/pdplgutsys
```

- Windows の場合

```
%PDDIR%\plugin\_phcfwd\conf\pdplgutsys
```

サーバ単位

各サーバ単位に定義情報を設定するユーザは、プラグインサーバ環境情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

HiRDB/ パラレルサーバの場合

- UNIX の場合

```
$PDDIR/plugin/_phcfwd/conf/BES 名称
```

- Windows の場合

```
%PDDIR%\plugin\_phcfwd\conf\BES 名称
```

HiRDB/ シングルサーバの場合

- UNIX の場合

```
$PDDIR/plugin/_phcfwd/conf/SDS 名称
```

- Windows の場合

```
%PDDIR%\plugin¥_phcfd¥conf¥SDS 名称
```

優先順位は、サーバ単位 > ユニット単位 > システム単位です。

記述誤りなどが検出された場合、メッセージログに警告メッセージを出力し、誤りのある定義ファイルは無視されます。環境定義ファイルがないか、環境定義ファイルはあるが設定されていない定義がある場合は、エラーではなく、定義値なしとして処理します。最終的に有効な値がない場合は、システムのデフォルトで動作します。

FREEWORD プラグインの環境定義ファイルの定義例を次の図に示します。

図 2-57 FREEWORD プラグインの環境定義ファイルの定義例

```
###FREEWORDプラグイン環境定義ファイル
set phc_search_wait=nowait      #   登録・検索同時実行指定      #
set phc_errorcode_detail=off    #   エラーコード詳細化指定      #
```

注意事項

- 各設定項目の行末は改行で区切ってください。
- 設定項目の途中を改行で区切らないでください。
- 1 行に設定できる文字列長は 1,023 バイトです。
- # (シャープ) 以降は、コメント文として扱われます。
- 各行は、空行なしで詰めて記述してください。

(2) 設定できる項目

FREEWORD プラグインを動作させる環境情報を FREEWORD プラグインの環境定義ファイルに定義することで、ユーザごとの環境を設定できます。設定できる定義項目を次の図に示します。なお、設定値項目には、ユーザごとに、必要に応じて取得できる最大値を設定します。

表 2-13 FREEWORD プラグインの環境定義ファイルに設定できる定義項目

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を定義できる単位 |
|-----------|-------------|-----------------------------|--|-----------------|
| 登録・検索同時実行 | 登録・検索同時実行指定 | set phc_search_wait | wait, nowait, または nowait_with_update デフォルト: wait | サーバ |
| トラブルシュート | エラーコード詳細化指定 | set phc_errorcode_detail | off または on デフォルト: off | システム, ユニット, サーバ |

登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行指定

登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行を使用するかどうか

かを指定します。

なお、この項目を設定できる単位は、サーバ単位だけです。

形式

```
set phc_search_wait=wait | nowait | nowait_with_update
```

- wait
同時実行を使用しない。
- nowait
登録・検索同時実行を使用する。
- nowait_with_update
登録・検索同時実行機能に加えて、分割遅延登録・更新系 SQL 同時実行を使用する。

エラーコード詳細化指定

SQL 連絡領域にメッセージが返却される場合に、エラーコードを詳細モードで出力するかどうかを指定します。

詳細モードで出力しない場合、SQLCODE=-997 のエラーとなります。

詳細モードで出力する場合、Text Search Plugin で出力したエラーコードが出力されます。エラーコードの詳細は「付録 F.1 SQL 連絡領域に出力されるメッセージ」を参照してください。

形式

```
set phc_errorcode_detail=off | on
```

- off
詳細化モードで出力しません。
- on
詳細化モードで出力します。

(3) 有効範囲

環境定義ファイルに定義した環境情報は、サーバ開始から終了まで有効です。HiRDB 稼働中に環境情報を変更した場合は、正しく動作しない場合がありますので、変更しないでください。

2.8.2 IXFREEWORD プラグインの環境設定

(1) 環境情報の定義方法

IXFREEWORD プラグインの環境情報は、次の単位で設定できます。IXFREEWORD プラグインの環境定義ファイルは、サンプルファイルとして提供されています。なお、HiRDB/ シングルサーバの場合は、次のどの単位で設定しても同じ意味になります。

システム単位

システム単位に定義情報を設定するユーザは、プラグインシステム定義情報ファイ

ルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phxfwr/conf/pdplgsys
```

- Windows の場合

```
%PDDIR%\plugin\_phxfwr\conf\pdplgsys
```

ユニット単位

ユニット単位に定義情報を設定するユーザは、プラグインユニット定義情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phxfwr/conf/pdplgutsys
```

- Windows の場合

```
%PDDIR%\plugin\_phxfwr\conf\pdplgutsys
```

サーバ単位

各サーバ単位に異なる定義情報を設定するユーザは、プラグインサーバ環境情報ファイルを作成します。

作成するディレクトリおよびファイル名は次のとおりです。

- UNIX の場合

```
$PDDIR/plugin/_phxfwr/conf/BES 名称
```

- Windows の場合

```
%PDDIR%\plugin\_phxfwr\conf\BES 名称
```

優先順位は、サーバ単位 > ユニット単位 > システム単位です。

記述誤りなどが検出された場合、メッセージログに警告メッセージを出力し、誤りのある定義ファイルは無視されます。環境定義ファイルがないか、環境定義ファイルはあるが設定されていない定義がある場合は、エラーではなく、定義値なしとして処理します。最終的に有効な値がない場合は、システムのデフォルトで動作します。

IXFREEWORD プラグインの環境定義ファイルの定義例を次の図に示します。

図 2-58 IXFREEWORD プラグインの環境定義ファイルの定義例

```

###IXFREEWORDプラグイン環境定義ファイル
set phx_mem_build=xxxx # 一括登録時の最大目盛りサイズ #
set phx_mem_insert=xxxx # 1件登録時の最大メモリサイズ #
set phx_mem_search=xxxx # 検索時最大ワークメモリサイズ #
set phx_dir_search=xxxxxxxxxx # 検索用ワークファイルディレクトリ #
set phx_opt_search=memory # 検索オプション #
set phx_opt_score=boolean # スコア算法指定 #
set phx_mem_expand=xxxx # 同義語異表記展開用最大メモリサイズ #
set phx_increment_practical_use=off # 異表記展開検索時の部分展開検索モード #
set phx_lock_pool_size=xxxx # 1RDエラー当たりの排他制御用共用メモリサイズ #
set phx_index_search=xxx # 検索抑止インデクスサイズ #
set phx_igram_proximity_search=on # 1文字近傍検索実行可否 #
set phx_delay_over=continue # 分割遅延登録用ファイルオーバ #
set phx_global_buffer_mode=0 # インデクス用グローバルバッファ最適化モード指定 #
    
```

注意事項

- 各設定項目の行末は改行で区切ってください。
- 1行に1項目だけ指定できます。2項目以上指定した場合は、2項目目以降は無視されます。
- 設定項目の途中を改行で区切らないでください。
- 検索用ワークファイルを使用して実行した検索が異常終了した場合、検索用ワークファイルディレクトリに不要なファイルが残ることがあります。これらのファイルは定期的にユーザが削除してください。
- 1行に設定できる文字列長は1,023バイトです。
- # (シャープ)以降は、コメント文として扱われます。
- 各行は、空行なしで詰めて記述してください。

(2) 設定できる項目

IXFREEWORD プラグインを動作させる環境情報を IXFREEWORD プラグインの環境定義ファイルに定義することで、ユーザごとの環境を設定できます。設定できる定義項目を次の表に示します。なお、各設定値項目には、ユーザごとに、必要に応じて取得できる最大値を設定します。

表 2-14 IXFREEWORD プラグインの環境定義ファイルに設定できる定義項目

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を定義できる単位 |
|----|-----------------------|-------------------|---|-----------------|
| 登録 | 一括登録時 最大メモリ サイズ | set phx_mem_build | 最大値：2,097,152 キロバイト 最小値：30,720 キロバイト デフォルト：51,200 キロバイト | システム, ユニット, サーバ |

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を定義できる単位 |
|------|----------------------------------|------------------------------------|--|-----------------|
| | 1 件登録時 最大メモリ サイズ | set phx_mem_insert | 最大値：2,097,152 キロバイト 最小値：3,072 キロバイト デフォルト：51,200 キロバイト | システム, ユニット, サーバ |
| 全文検索 | 検索時最大 ワークメモリ サイズ | set phx_mem_search | 最大値：2,097,152 キロバイト 最小値：2,048 キロバイト デフォルト：20,480 キロバイト | システム, ユニット, サーバ |
| | 検索用ワーク ファイル ディレクトリ | set phx_dir_search | 最大長：200 バイト (¥0 を含む) 最小長：1 バイト UNIX の場合のデフォルト：HiRDB 運用ディレクトリ /plugin/_phxfwr/tmp Windows の場合のデフォルト：HiRDB 運用フォルダ ¥plugin¥_phxfwr¥tmp | システム, ユニット, サーバ |
| | 検索オプション | set phx_opt_search | memory または file デフォルト：memory | システム, ユニット, サーバ |
| | スコア算出 法指定 | set phx_opt_score | boolean または total デフォルト：boolean | システム, ユニット, サーバ |
| | 同義語異表 記展開用最大 メモリサイズ | set phx_mem_expand | 最大値：2,097,152 キロバイト 最小値：1 キロバイト デフォルト：0 (無制限) | システム, ユニット, サーバ |
| | 異表記展開 検索時の部分 展開検索モード 指定 | set phx_increment_practical_use | on または off デフォルト：off | システム, ユニット, サーバ |
| | 検索抑止イン デックスサイズ | set phx_index_search | 最大値：2,097,152 キロバイト 最小値：64 キロバイト デフォルト：2,097,152 キロバイト | システム, ユニット, サーバ |
| | 1 文字近傍 検索実行可否 | set phx_lgram_proximity_search | on または off デフォルト：on | システム, ユニット, サーバ |

2. Text Search Plug-in システムの構築

| 種別 | 環境設定項目 | パラメタ | 設定値 | 環境を定義できる単位 |
|---------------|--|-----------------------------------|---|-------------------|
| 登録・検索 同時実行 | 1RD エリア 当たりの排 他制御用共 用メモリサ イズ | set phx_lock_pool_size | 最大値：1,024 キロバ イト 最小値：8 キロバイト デフォルト：8 キロバ イト | サーバ |
| 分割遅延 登録 | 分割遅延登 録用ファ イルオーバ時 の処理 | set phx_delay_over | continue または error デフォルト：continue | システム，ユニット， サーバ |
| 運用 | インデクス 用グローバ ルバッファ 最適化モ ード指定 | set phx_global_buffer_ mode | 0 または 1 デフォルト：0 | システム，ユニット， サーバ |

一括登録時最大メモリサイズ

IXFREEWORD インデクスに対する一括追加 (CREATE INDEX, pdload) をするときに使用する最大メモリサイズ (キロバイト) です。

形式

```
set phx_mem_build=xxxx
```

1 件登録時最大メモリサイズ

IXFREEWORD インデクスに 1 件追加 (INSERT 文, UPDATE 文, pdload の SYNC モード) するとき使用する最大メモリサイズ (キロバイト) です。

形式

```
set phx_mem_insert=xxxx
```

検索時最大ワークメモリサイズ

検索処理をするときに使用する作業用メモリの最大メモリサイズ (キロバイト) です。

形式

```
set phx_mem_search=xxxx
```

検索用ワークファイルディレクトリ

検索オプションに, "set phx_opt_search=file" を指定した場合, 検索に使用するワークファイルを作成するディレクトリ名を絶対パスで指定します。Text Search Plug-in は, 検索開始時にファイル名が "swk" で始まるユニークなファイルを指定されたディレクトリに作成し, 検索終了時に削除します。

検索が複数同時実行された場合, ディレクトリ競合による性能低下のおそれがあります。そのため, 同一の BES に対して複数のユーザが同時に検索する運用の場合は, ファイルではなくメモリの使用をお勧めします。検索処理が異常終了した場合は, 作成されたワークファイルは削除されないため, HiRDB 管理者が削除してください。

形式

```
set phx_dir_search=xxxxxxxxxxxxxx
```

検索オプション

検索時の動作方法を指定します。

形式

```
set phx_opt_search=memory | file
```

- memory
メモリ上で動作します。set phx_mem_search で指定したメモリサイズを超える場合は、メモリ不足エラーとなります。
- file
メモリおよび検索用ワークファイルで動作します。

同義語異表記展開用最大メモリサイズ

同義語異表記展開処理をするときに使用する最大メモリサイズ(キロバイト)です。0を指定した場合は、無制限となります。

同義語異表記展開用メモリサイズの見積もりについては、「2.3 リソースの見積もり」を参照してください。

なお、本パラメタは異表記展開検索時の部分展開検索モードが"on"(全展開検索モード)の場合に有効となります。

形式

```
set phx_mem_expand=xxxx
```

異表記展開検索時の部分展開検索モード

異表記展開検索時に、展開処理を部分的に実行するか、全面展開するかを指定します。この指定は、環境がシステム、ユニット、またはサーバの場合に有効です。

形式

```
set phx_increment_practical_use = on | off
```

- on
全展開検索モードで検索します。
- off
部分展開検索モードで検索します。

かたかな異表記展開検索を実行する場合は、部分展開検索モード指定(off指定)を推奨します。

スコア算出法指定

検索時のスコア算出法を指定します。

形式

```
set phx_opt_score = boolean | total
```

- boolean
論理演算時に、検索条件が論理積の場合は演算対象内で最小のスコアを最終スコアとし、論理和の場合は演算対象内で最大のスコアを最終スコアとし

2. Text Search Plug-in システムの構築

す。

- total

論理演算時に、検索条件が論理積、論理和のどちらの場合でも、演算対象となるスコアの総和の値を最終スコアとします。

検索抑止インデクスサイズ

検索で使用するインデクスサイズの上限值を指定します。検索時に指定した上限値を超えるインデクスを参照すると、SQL 連絡領域に次のメッセージが出力されます。

```
Index data for search exceeds maximum size.(XX)
```

メッセージの詳細については、「付録 F エラーメッセージ」を参照してください。

形式

```
set phx_index_search = xxx
```

1 文字近傍検索実行可否

1 文字のインデクスを使用する近傍検索や、"A*A" のような 1 文字のインデクスを使用する前方後方一致のワイルドカード検索の実行を可能にするかを指定します。なお、これらの検索は、参照するインデクスが多くなり、システムに負荷を与えることがあります。このため、これらの検索を実行する必要がない場合は「off」を指定してください。

形式

```
set phx_lgram_proximity_search = on | off
```

- on

1 文字のインデクスを使用する近傍検索や前方後方一致のワイルドカード検索を実行できます。

- off

1 文字のインデクスを使用する近傍検索や前方後方一致のワイルドカード検索を実行するとエラーリターンして、SQL 連絡領域に次のメッセージが出力されます。

```
Cannot specify term of one character in proximity.
```

メッセージの詳細については、「付録 F エラーメッセージ」を参照してください。

1RD エリア当たりの排他制御用共用メモリサイズ

登録・検索同時実行時の排他制御に使用する共用メモリのサイズ(キロバイト)を指定します。サイズは 8 キロバイトの倍数で指定してください。8 キロバイトの倍数でない値が指定された場合は、8 キロバイト単位で切り下げられます。ただし、8 キロバイトより小さい値が指定された場合は、8 キロバイト(デフォルト)になります。

なお、この項目を設定できる単位は、サーバ単位だけです。

形式

```
set phx_lock_pool_size=xxxx
```

分割遅延登録用ファイルオーバ時の処理

インデックス分割遅延登録時に、指定された分割登録用の領域を使い切った場合の処理を指定します。

形式

```
set phx_delay_over=continue | error
```

- continue
登録要求された文書は遅延せず、即時にインデックスを作成して登録します。
- error
エラーリターンします。

インデックス用グローバルバッファ最適化モード指定

インデックスにアクセスするときの HiRDB BLOB 用グローバルバッファを最適化するかどうか指定します。最適化した場合、検索性能の向上が期待できます。

形式

```
set phx_global_buffer_mode = 0 | 1
```

- 0
最適化しません。
- 1
最適化します。
検索用の管理データをファイル単位で優先してバッファリングします。各ファイルの優先付けについては、「5.2.1 グローバルバッファ管理の運用」を参照してください。

(3) 有効範囲

環境定義ファイルに定義した環境情報は、サーバ開始から終了まで有効です。HiRDB稼働中に定義を変更した場合は、正しく動作しない場合がありますので、変更しないでください。なお、シングルサーバ構成では、シングルサーバ = BES となります。

2.9 システムの制限

Text Search Plug-in のシステムの制限について説明します。

Text Search Plug-in では、インデクス情報格納用 RD エリアサイズの上限值、および SGMLTEXT 型列、FREEWORD 型列を含んだ表について、次の制限があります。

インデクス情報格納用 RD エリアサイズの上限值

インデクスファイル自動拡張の場合：32 ギガバイト

インデクスファイル自動拡張しない場合：64 ギガバイト

SGMLTEXT 型列の 1 列あたりの最大登録件数 ¹

- 表分割を行わない場合
800 万件
- 表分割を行う場合の 1 分割表の各列あたりの最大登録件数
800 万件

FREEWORD 型列の 1 列あたりの最大登録件数 ¹

- 表分割を行わない場合
1600 万件 ²
- 表分割を行う場合の 1 分割表の各列あたりの最大登録件数
1600 万件 ²

注 1

最大登録件数は削除済みのデータを含んだ件数です。
削除データを無効化するには、データベース再編成ユーティリティ (pdrrorg) による再編成、またはインデクスの再作成を実行してください。

注 2

800 万件を超える場合、CREATE INDEX のプラグインオプションで登録件数指定 (DOC_COUNT) を行ってください。

2.10 Text Search Plug-in の設定解除

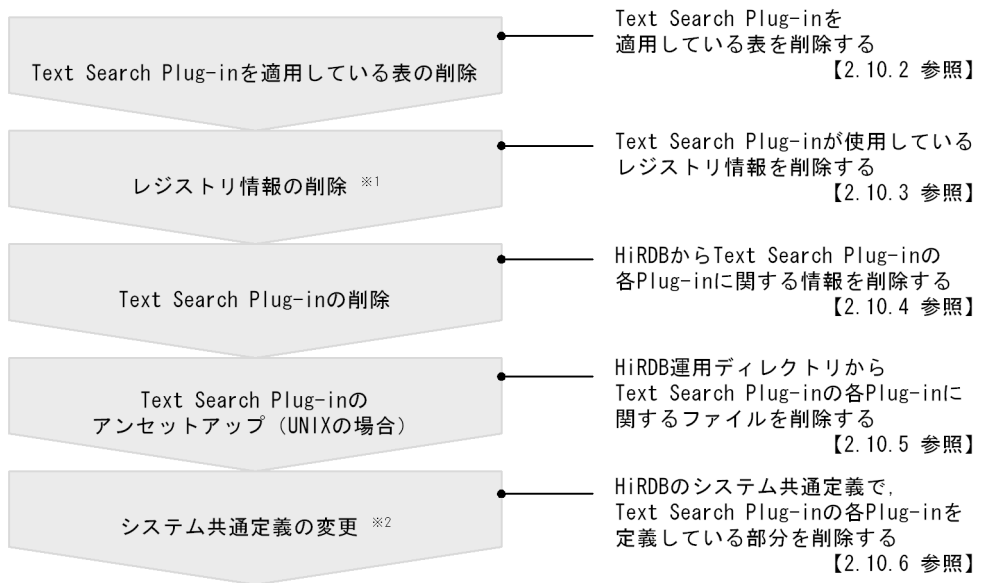
この節では、コマンドを使用した、Text Search Plug-in の設定解除方法について説明します。ここでは、文書検索プラグインまたは文字列検索プラグインのデータベースを運用していることを前提に説明します。

なお、HiRDB が提供しているユティリティの使用方法については、マニュアル「HiRDB コマンドリファレンス」および「HiRDB システム導入・設計ガイド」を参照してください。また、Text Search Plug-in のユティリティの使用方法については、「3. Text Search Plug-in のユティリティ」を参照してください。

2.10.1 設定解除の流れ

文書検索プラグインまたは文字列検索プラグインのデータベースの運用状態から、Text Search Plug-in の設定解除までの操作の流れを次の図に示します。

図 2-59 Text Search Plug-in の設定解除の流れ



注※1 SGMLTEXT型の列を定義していた場合だけに必要な作業です。

注※2 登録・検索同時実行を指定した場合は、HiRDBのバックエンドサーバ定義またはシングルサーバ定義に追加した、Plug-in定義を削除する必要があります。

以降、上記の流れに従って作業の詳細を説明します。

2.10.2 表の削除

Text Search Plug-in を適用している表を削除するには、定義系 SQL の DROP TABLE を使用します。SGMLTEXT 型または FREEWORD 型の列を定義している表が削除の対象となります。DROP TABLE の詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

2.10.3 レジストリ情報の削除（SGMLTEXT 型の列を定義していた場合）

SGMLTEXT 型の列を定義していた場合は、Text Search Plug-in が使用しているレジストリ情報を削除します。Text Search Plug-in が使用しているレジストリ情報の削除には、「phsregsetup（レジストリ初期化ユティリティ）」を使用します。なお、FREEWORD 型の列だけを定義していた場合は、レジストリ情報の削除は不要です。

レジストリ初期化ユティリティの指定例を次に示します。

```
phsregsetup -d
```

2.10.4 Text Search Plug-in の削除

Text Search Plug-in の削除には、HiRDB の運用コマンド「pdplgrgst（プラグインの登録・削除）」を使用します。

Text Search Plug-in の削除には、HiRDB の運用コマンド「pdplgrgst（プラグインの登録・削除）」を使用します。プラグインを削除する場合、pdplgrgst コマンド実行前に次のものを削除しておいてください。

- 削除するプラグインを使用した表
- 削除するプラグインを使用したインデクス
- 削除するプラグインが提供している抽象データ型、インデクス型、および関数を利用した、関数、手続き

Text Search Plug-in を削除する場合は、次の形式で実行してください。なお、文書検索プラグインを削除する場合には、n-gram インデクスプラグイン、SGML プラグインの順に削除してください。また、文字列検索プラグインを削除する場合には、IXFREEWORD プラグイン、FREEWORD プラグインの順に削除してください。

コマンドの形式

```
pdplgrgst -d プラグイン定義ファイル名 PICファイル名
```

引数

プラグイン定義ファイル名

プラグイン定義ファイルのファイル名を指定します。コマンドを入力するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名

を指定します。Text Search Plug-in では、各プラグインに対応してプラグイン定義ファイル名を次のように指定してください。

| 文書検索プラグイン | プラグイン定義ファイル名 |
|-------------------|--------------|
| SGML プラグイン | _phsgml.adt |
| n-gram インデクスプラグイン | _phngram.idx |

| 文字列検索プラグイン | プラグイン定義ファイル名 |
|-----------------|--------------|
| FREWORD プラグイン | _phcfwd.adt |
| IXFREWORD プラグイン | _phxfwr.idx |

PIC ファイル名

PIC ファイルのファイル名を指定します。コマンドを入力するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定します。Text Search Plug-in では、各プラグインに対応して PIC ファイル名を次のように指定してください。

| 文書検索プラグイン | PIC ファイル名 |
|-------------------|--------------|
| SGML プラグイン | _phsgml.pic |
| n-gram インデクスプラグイン | _phngram.pic |

| 文字列検索プラグイン | PIC ファイル名 |
|-----------------|-------------|
| FREWORD プラグイン | _phcfwd.pic |
| IXFREWORD プラグイン | _phxfwr.pic |

コマンド実行例 (UNIX の場合)

```
pdplgrgst -d $PDDIR/plugin/_phngram/etc/_phngram.idx ¥
           $PDDIR/plugin/_phngram/etc/_phngram.pic
pdplgrgst -d $PDDIR/plugin/_phsgml/etc/_phsgml.adt ¥
           $PDDIR/plugin/_phsgml/etc/_phsgml.pic
pdplgrgst -d $PDDIR/plugin/_phxfwr/etc/_phxfwr.idx ¥
           $PDDIR/plugin/_phxfwr/etc/_phxfwr.pic
pdplgrgst -d $PDDIR/plugin/_phcfwd/etc/_phcfwd.adt ¥
           $PDDIR/plugin/_phcfwd/etc/_phcfwd.pic
```

上記は、2行でコマンドを入力した例を記載していますが、1行で入力する場合は、"¥" (エスケープ文字) を入力しないでください。

コマンド実行例 (Windows の場合)

2. Text Search Plug-in システムの構築

```
pdplgrgst -d %PDDIR%¥plugin¥_phngram¥etc¥_phngram.idx
           %PDDIR%¥plugin¥_phngram¥etc¥_phngram.pic
pdplgrgst -d %PDDIR%¥plugin¥_phsgml¥etc¥_phsgml.adt
           %PDDIR%¥plugin¥_phsgml¥etc¥_phsgml.pic
pdplgrgst -d %PDDIR%¥plugin¥_phxfwr¥etc¥_phxfwr.idx
           %PDDIR%¥plugin¥_phxfwr¥etc¥_phxfwr.pic
pdplgrgst -d %PDDIR%¥plugin¥_phcfwd¥etc¥_phcfwd.adt
           %PDDIR%¥plugin¥_phcfwd¥etc¥_phcfwd.pic
```

- 「PDDIR」は、HiRDB インストール先フォルダパスが格納された環境変数を示します。環境変数 PDDIR は「%」で囲んで「%PDDIR%」と入力してください。
- 上記の例では、コマンドの引数を改行して記述していますが、実際は改行しないで入力してください。

2.10.5 Text Search Plug-in のアンセットアップ (UNIX の場合)

HiRDB 運用ディレクトリから Text Search Plug-in の各プラグインに関するファイルを削除します。Text Search Plug-in のアンセットアップには、HiRDB の運用コマンド「pdplgset (プラグインのセットアップ)」を使用します。HiRDB の運用コマンド「pdstop (HiRDB システム・ユニット・サーバの終了)」で正常終了させたあとに実行してください。同義語辞書のバックアップが必要な場合は、Text Search Plug-in のアンセットアップをする前に、同義語辞書をバックアップしてください。バックアップを実行するときには、「2.10.7 HiRDB を OS から削除する場合の注意事項」、および「5.3.2 バックアップ取得方法」を参照してください。

Text Search Plug-in のアンセットアップは、次の形式で実行してください。

コマンドの形式

```
pdplgset -d プラグイン名
```

引数

-d プラグイン名

アンセットアップするプラグインの名称を指定します。Text Search Plug-in では、各プラグインに対応してプラグイン名を次のように指定してください。

| 文書検索プラグイン | プラグイン名 |
|-------------------|----------|
| SGML プラグイン | _phsgml |
| n-gram インデクスプラグイン | _phngram |

| 文字列検索プラグイン | プラグイン名 |
|---------------|---------|
| FREWORD プラグイン | _phcfwd |

| 文字列検索プラグイン | プラグイン名 |
|------------------|----------|
| IXFREEWORD プラグイン | _phxfwrд |

コマンド実行例

```
pdplgset -d _phsgml
pdplgset -d _phngram
pdplgset -d _phcfwd
pdplgset -d _phxfwrд
```

2.10.6 システム共通定義の変更

HiRDB のシステム共通定義「pdsys」に追加した、使用するプラグインを宣言する「pdplugin」オペランドで Text Search Plug-in の各プラグインを定義している部分を削除します。このとき、このオペランドを追加しているすべてのサーバマシン上のシステム共通定義から削除してください。なお、この作業は、HiRDB の運用コマンド「pdstop (HiRDB システム・ユニット・サーバの終了)」で正常終了させたあとに実行してください。

削除するオペランドの例を次に示します。

```
pdplugin -n _phsgml
pdplugin -n _phngram
pdplugin -n _phcfwd
pdplugin -n _phxfwrд
```

また、登録・検索同時実行を使用していた場合は、HiRDB のバックエンドサーバ定義またはシングルサーバ定義に追加した、プラグインで使用する共用メモリのサイズを宣言する「pdplgprm」オペランドで、n-gram インデックスプラグインまたは FREEWORD プラグインで使用する共用メモリのサイズを定義している部分を削除します。このとき、このオペランドを追加しているすべてのサーバマシン上のバックエンドサーバ定義またはシングルサーバ定義から削除してください。

削除するオペランドの例を次に示します。

```
pdplgprm -n _phngram -s 128
pdplgprm -n _phxfwrд -s 128
```

2.10.7 HiRDB を OS から削除する場合の注意事項 (UNIX の場合)

HiRDB の運用コマンド「pdsetup -d (HiRDB システムの OS からの削除)」を実行して HiRDB を OS から削除する場合は、まず HiRDB の運用コマンド「pdplgset -d (プラグインのアンセットアップ)」を実行して、Text Search Plug-in を先にアンセットアップしてください。

ただし、「pdplgset -d」を実行するとプラグイン固有ファイルが削除されるため、バックアップが必要な場合はアンインストールの前にバックアップを実行してください。バツ

2. Text Search Plug-in システムの構築

クアップの詳細については、「5.3.2 バックアップ取得方法」を参照してください。

2.11 HiRDB の提供する機能の利用

Text Search Plug-in は、HiRDB に組み込んで利用します。このため、HiRDB の次の機能を利用できます。

データベース作成ユティリティ (pload)

HiRDB の「pload (データベース作成ユティリティ)」を使用して、Text Search Plug-in にデータを登録できます。

pload を使用して Text Search Plug-in にデータを登録するときには、列構成情報ファイルにコンストラクタ関数を記述する必要があります。

コンストラクタ関数の情報を次の表に示します。

表 2-15 コンストラクタ関数の情報

| 抽象データ型 | コンストラクタ関数名 (func) | 引数の型 (param) |
|------------|---------------------|----------------|
| SGMLTEXT 型 | SGMLTEXT | blob |
| FREWORD 型 | FREWORD | varchar |

定義ファイルでの抽象データ型関数の指定例を次に示します。指定方法の詳細については、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。

SGMLTEXT 型の場合

```
doc, func=(SGMLTEXT,param=blob)
```

FREWORD 型の場合

```
doc, func=(FREWORD,param=varchar)
```

データベース再編成ユティリティ (pdrorg)

Text Search Plug-in では、HiRDB の「pdrorg (データベース再編成ユティリティ)」を使用した表の再編成またはインデクスの一括作成に対応しています。

Text Search Plug-in では、pdrorg に対応した次の機能を提供しています。

- アンロード機能
- プラグインインデクス遅延一括作成機能
- プラグインインデクス一括作成部分回復機能

pdrorg を使用する場合、コンストラクタ関数およびコンストラクタパラメタ逆生成関数の指定が必要になることがあります。コンストラクタパラメタ逆生成関数の情報を次の表に示します。

表 2-16 コンストラクタパラメタ逆生成関数

| 抽象データ型 | コンストラクタパラメタ逆生成関数 (func) | 引数の型 (param) |
|------------|---------------------------|----------------|
| SGMLTEXT 型 | unsgmltext | sgmltext |

2. Text Search Plug-in システムの構築

| 抽象データ型 | コンストラクタパラメタ逆生成関数 (func) | 引数の型 (param) |
|------------|-------------------------|--------------|
| FREEWORD 型 | extracts | freeword |

定義ファイルでの抽象データ型関数の指定例を次に示します。指定方法の詳細については、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。

SGMLTEXT 型の場合

アンロード時

```
unld_func type=SGMLTEXT,func=unsgmltext (sgmltext)
```

リロード時

```
reld_func type=SGMLTEXT,func=sgmltext (blob)
```

FREEWORD 型の場合

アンロード時

```
unld_func type=FREEWORD,func=extracts (freeword)
```

リロード時

```
reld_func type=FREEWORD,func=freeword (varchar)
```

絞込み検索でのリスト機能拡張 (全文検索でのスコア値取得)

Text Search Plug-in では、HiRDB の「絞込み検索でのリスト機能拡張 (全文検索でのスコア値取得)」を使用したリストでの絞込み検索に対応しています。指定方法の詳細については、マニュアル「HiRDB UAP 開発ガイド」「プラグイン提供関数の制限」を参照してください。

プラグイン提供関数からの一括取得処理

Text Search Plug-in では、HiRDB の「プラグイン提供関数からの一括取得処理」および「プラグイン提供関数からの返却結果の一括取得処理の適用範囲拡大」に対応しています。指定方法の詳細については、マニュアル「HiRDB UAP 開発ガイド」「クライアント環境定義の設定内容」の環境変数「PDSQLOPTLVL」の「プラグイン提供関数からの一括取得機能」を参照してください。

3

Text Search Plug-in のユティリティ

この章では、Text Search Plug-in の運用に必要なユティリティの使用方法について説明します。

3.1 ユティリティの使用方法

3.2 文書検索プラグインのユティリティ

3.3 文字列検索プラグインのユティリティ

3.1 ユティリティの使用方法

この節では、Text Search Plug-in が提供するユティリティの使用方法について説明します。

3.1.1 Text Search Plug-in のユティリティとは

Text Search Plug-in では、Text Search Plug-in の設定、管理および運用のためにユティリティを提供しています。ユティリティは、Text Search Plug-in を登録したマシン上で実行します。

なお、Text Search Plug-in の運用中には、Text Search Plug-in のユティリティ以外に、HiRDB で提供されているユティリティや運用コマンドを使用する場合があります。HiRDB で提供されているユティリティおよび運用コマンドの使用方法については、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。

Text Search Plug-in が提供する文書検索プラグインおよび文字列検索プラグインのユティリティの機能一覧を、表 3-1 に示します。

表 3-1 Text Search Plug-in のユティリティの機能一覧

| プラグイン種別 | ユティリティコマンド名 | 機能 | ユティリティコマンドの実行者 | HiRDB稼働中での実行可否 | 実行する HiRDB 運用環境 | |
|-----------|-------------|--------------|----------------|----------------|-----------------|---|
| 文書検索プラグイン | SGML プラグイン | phsregsetup | レジストリの初期化 | HiRDB 管理者 | | SDS または MGR |
| | | phssgmlreg | SGML 定義情報の登録 | HiRDB 管理者 | | SDS または MGR |
| | | phssgmlrm | SGML 定義情報の削除 | HiRDB 管理者 | | SDS または MGR |
| | | phssgmlval | SGML 定義情報の参照 | HiRDB 管理者 | | SDS または MGR |
| | | phssyndicreg | 同義語辞書の登録 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ¹ |
| | | phssyndicrm | 同義語辞書の削除 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ¹ |

| プラグイン種別 | ユーティリティコマンド名 | 機能 | ユーティリティコマンドの実行者 | HiRDB稼働中での実行可否 | 実行する HiRDB 運用環境 |
|-------------------|--------------------------|-----------------------------|-----------------|----------------|---|
| | phssyndicconv | 同義語辞書の移行 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ¹ |
| | phssyndicbkup (UNIX の場合) | 同義語辞書のバックアップ | HiRDB 管理者 | × | Text Search Plug-in を登録したすべての HiRDB 運用環境 ¹ |
| n-gram インデクスプラグイン | phnidxls | インデクス情報の取得 | HiRDB 管理者 | | SDS または MGR |
| | phninemt ³ | インデクスのインクリメンタル (最適化) | HiRDB 管理者 | | SDS または MGR |
| | phnmerge | 差分インデクスから MASTER インデクスへのマージ | HiRDB 管理者 | | SDS または MGR |
| | phnexpnd | インデクスの拡張 | HiRDB 管理者 | | SDS または MGR |
| | phnstartidx | インデクスの作成開始 | HiRDB 管理者 | | SDS または MGR |
| | phndelay | インデクス遅延更新開始 / 終了指定 | HiRDB 管理者 | | SDS または MGR |
| | phnmodidx | インデクス情報変更 | HiRDB 管理者 | | SDS または MGR |
| | phncond | インデクス詰め替え | HiRDB 管理者 | | SDS または MGR |
| | phnconget ³ | 統計情報収集 | HiRDB 管理者 | | SDS または MGR |
| | phnconbuild ³ | 統計情報統合 | HiRDB 管理者 | | SDS または MGR |
| | phnconstr ³ | 統計情報反映 | HiRDB 管理者 | | SDS または MGR |

3. Text Search Plug-in のユーティリティ

| プラグイン種別 | ユーティリティコマンド名 | 機能 | ユーティリティコマンドの実行者 | HiRDB稼働中での実行可否 | 実行する HiRDB 運用環境 | |
|------------|------------------------|--------------------------|----------------------|----------------|-----------------|---|
| | phnchgcon ³ | 特徴ターム取得 RD エリア指定 | HiRDB 管理者 | | SDS または MGR | |
| 文字列検索プラグイン | FREE WORD プラグイン | phcsyndicreg | 同義語辞書の登録 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ² |
| | | phcsyndicrm | 同義語辞書の削除 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ² |
| | | phcsyndicconv | 同義語辞書の移行 | HiRDB 管理者 | | Text Search Plug-in を登録したすべての HiRDB 運用環境 ² |
| | | phcsyndicbkup (UNIX の場合) | 同義語辞書のバックアップ | HiRDB 管理者 | × | Text Search Plug-in を登録したすべての HiRDB 運用環境 ² |
| | IXFREE WORD プラグイン | phxidxls | インデクス情報の取得 | HiRDB 管理者 | | SDS または MGR |
| | | phxincmt ³ | インデクスのインクリメンタル (最適化) | HiRDB 管理者 | | SDS または MGR |
| | | phxexpnd | インデクスの拡張 | HiRDB 管理者 | | SDS または MGR |
| | | phxstartidx | インデクスの作成開始 | HiRDB 管理者 | | SDS または MGR |
| | | phxdelay | インデクス遅延更新開始 / 終了指定 | HiRDB 管理者 | | SDS または MGR |
| | | phxmodidx | インデクス情報変更 | HiRDB 管理者 | | SDS または MGR |
| phxcond | インデクス詰め替え | HiRDB 管理者 | | SDS または MGR | | |

(凡例)

- : HiRDB が稼働中のときだけ実行できます。
- × : HiRDB が停止しているときだけ実行できます。
- SDS : HiRDB/ シングルサーバの場合に、シングルサーバがあるサーバマシンで実行することを示します。
- MGR : HiRDB/ パラレルサーバの場合に、システムマネージャがあるサーバマシンで実行することを示します。

注

ユティリティの実行時に指定したファイルを、ほかのユティリティまたは UAP で同時に使用しないでください。同じファイルをユティリティまたは UAP で同時に使用した場合は、処理結果が誤りとなったり、エラーが発生したりすることがあります。

注 1

SGML プラグインを登録したすべての HiRDB 運用環境に対して、同一の操作をしてください。

注 2

FREEWORD プラグインを登録したすべての HiRDB 運用環境に対して、同一の操作をしてください。

注 3

英文インデクスは指定できません。

3.1.2 ユティリティの記述形式と操作

ここでは、ユティリティの入力方法と記述形式について説明します。なお、ユティリティは、open 状態かつ HiRDB の運用コマンド「pdhold (RD エリアの閉塞)」による閉塞状態か、open 状態かつ閉塞解除状態のどちらかのときに実行してください。

(1) ユティリティの記述形式

ユティリティは次の形式で記述します。

ユティリティコマンド名 オプション コマンド引数

(a) ユティリティコマンド名

ユティリティコマンド名は、実行するユティリティコマンドのファイル名です。各ユティリティコマンドを格納しているディレクトリ名については、「付録 B ディレクトリおよびファイル構成」を参照してください。

(b) オプション

オプションで指定するのは、ユティリティコマンドの操作そのものを修飾するか、または操作の対象物を特定するために、対象物を修飾するかのどちらかです。

次の説明中に使用する \$ はシェルのプロンプト、cmd はユティリティコマンド名を示します。

オプションは、- (マイナス記号) で始まる文字列で、フラグ引数を指定しないか、または 1 種類のフラグ引数を指定します。

3. Text Search Plug-in のユティリティ

オプションの記述形式を次に示します。

- オプションフラグ

または

- オプションフラグ フラグ引数

(凡例)

オプションフラグ：1文字以上の英数字（英大文字と英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

フラグ引数のないオプションフラグは、一つのマイナス記号の後ろにまとめて指定できません。

(例) 次の二つは同じ意味です。

```
$ cmd -a -b -c
```

```
$ cmd -abc
```

- 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。
- フラグ引数中にあるスペースを含める場合（UNIXの場合は、シェルから入力するとき）は、フラグ引数の前後を"（ダブルクォーテーション）で囲まなければなりません。

(例) 1 2 を引数に持つオプション f は、次のように記述します。

```
$ cmd -f "1 2"
```

同じオプションフラグを 2 回以上指定すると、最後に指定したオプションの値が有効になります。

(例) 次のように入力すると、-a 2 が有効となります。

```
$ cmd -a 1 -a 2
```

オプションは、コマンド引数よりも前に指定しなければなりません。

(例) オプションフラグ a が、フラグ引数をとらない場合、次のように入力すると、file と -b はコマンド引数とみなされます。

```
$ cmd -a file -b
```

二つのマイナス記号 "- -" は、オプションの終わりを示します。

(例) 次のように入力すると、-b はコマンド引数とみなされます。

```
$ cmd -a -- -b
```

マイナス記号だけのオプションは、入力できません。

(例) 次のように入力すると、- はコマンド引数とみなされます。

```
$ cmd -
```

(c) コマンド引数

コマンド引数は、ユティリティコマンドの操作で、直接対象になる値を指定します。対象物を修飾指定する場合、オプションで対象物を指定します。

コマンド引数として制御文ファイルを指定した場合、特に明記しないかぎり通常ファイルを対象とします。

3.2 文書検索プラグインのユティリティ

この節では、Text Search Plug-in の文書検索プラグイン（SGML プラグインおよび n-gram インデクスプラグイン）で、SGML 文書を扱うために必要なユティリティの詳細について説明します。

3.2.1 phsregsetup（レジストリ初期化ユティリティ）

（1）機能

Text Search Plug-in が使用する HiRDB のレジストリを初期化します。SGML 定義情報を登録する前に、レジストリを初期化してください。なお、-d オプションを付けた場合は、登録してあるレジストリ情報を削除します。

（2）実行者

HiRDB 管理者が実行できます。

（3）形式

```
phsregsetup [-d]
```

（4）コマンド引数

-d

Text Search Plug-in が使用しているレジストリ情報を削除する場合に指定します。

（5）規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

3.2.2 phssgmlreg (SGML 定義情報登録ユティリティ)

(1) 機能

SGML 文書を登録するために必要な SGML 定義情報を登録します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phssgmlreg 定義属性名 登録名 ローカルファイル名

(4) コマンド引数

定義属性名

登録する定義情報の種別を指定します。

- DTD : DTD
- NORparm : 正規化パラメタ

登録名

定義を登録する際に付ける名称を指定します。システム内で区別できるようなユニークな名称を、1 ~ 240 バイトの範囲内で指定してください。

ローカルファイル名

DTD および正規化パラメタを格納したローカルなファイルの名称を指定します。コマンドを入力するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定します。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

3.2.3 phssgmlrm (SGML 定義情報削除ユティリティ)

(1) 機能

phssgmlreg で登録した SGML 定義情報を削除します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phssgmlrm 定義属性名 登録名

(4) コマンド引数

定義属性名

削除する定義情報の種別を指定します。

- DTD : DTD
- NORparm : 正規化パラメタ

登録名

削除する SGML 定義情報の登録名を指定します。登録されている名称を指定してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

3.2.4 phssgmlval (SGML 定義情報参照ユティリティ)

(1) 機能

レジストりに登録した SGML 定義情報を参照します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phssgmlval 定義属性名 登録名 [出力ファイル名]

(4) コマンド引数

定義属性名

参照する定義情報の種別を指定します。

- DTD : DTD
- NORparm : 正規化パラメタ

登録名

参照したい SGML 定義情報の登録名を指定します。登録されている名称を指定してください。

出力ファイル名

取得した SGML 定義情報を出力するローカルファイルの名称を指定します。省略した場合は、標準出力に出力します。

ファイル名だけを指定した場合は、ユティリティを実行したときのカレントディレクトリに出力します。任意のディレクトリに出力させたい場合は、絶対パスを含めたファイル名を指定します。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

3.2.5 phssyndicreg (同義語辞書登録ユティリティ)

(1) 機能

同義語検索をするために必要な同義語辞書を登録します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phssyndicreg 同義語辞書名 同義語辞書テキストファイル名 [-u {jis | ms |  
jis_ucs4 | ms_ucs4}]
```

(4) コマンド引数

同義語辞書名

辞書を登録する際に付ける名称を指定します。システム内で区別できるユニークな名称を、1 ~ 32 バイトの範囲内で指定してください。

同義語辞書テキストファイル名

同義語辞書のテキストファイルの名称を指定します。
ユティリティを実行するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定します。

-u

Unicode 種別を指定します。データベースの文字コード種別が utf-8 の場合は、必ず指定します。データベースの文字コードが utf-8 でない場合には指定しないでください。

次のどれかを指定してください。

- jis : JISX0221 で UCS2 の範囲
- ms : MS-Unicode で UCS2 の範囲
- jis_ucs4 : JISX0221 で UCS4 の範囲
- ms_ucs4 : MS-Unicode で UCS4 の範囲

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

SGML プラグインを登録したすべての HiRDB 運用環境に対して、同一の同義語辞書テキストファイルを使用し、同一の同義語辞書名で登録してください。

文字列検索プラグインと同じ同義語辞書を使用する場合は、文字列検索プラグインで登録した同義語辞書テキストファイルを、文書検索プラグインにも同様に登録してください。

同義語辞書名に指定できる文字は、HiRDB をセットアップしたときの文字コード、記号英数字、および半角かたかなです。ただし、次の表に示す記号および英数字は指定できません。

データベースの文字コード種別が utf-8 の場合、UCS4 の範囲のインデクス (INDEX_UCS4) に対する同義語検索は、UCS4 の範囲で作成した同義語辞書を使用します。INDEX_UCS4 指定のないインデクスは、UCS2 の範囲で作成した同義語辞書を使用します。また、extract 関数は UCS4 の範囲の同義語辞書を優先的に使用しますが、UCS4 の範囲の同義語辞書が作成されていない場合には、UCS2 の範囲の同義語辞書を使用します。

系切り替え運用を行なう場合は、「付録 E 系切り替え」を参照してください。

表 3-2 同義語辞書名に指定できない記号および英数字

| 文字 | 文字の略称 |
|----|---------------|
| " | ダブルクォーテーション |
| , | コンマ |
| { | 左波括弧 |
| } | 右波括弧 |
| [| 左角括弧 |
|] | 右角括弧 |
| (| 左丸括弧 |
|) | 右丸括弧 |
| ¥ | 円記号 (エスケープ文字) |
| * | アスタリスク |
| ? | 疑問符 |
| | パイプ |
| ^ | ハット |
| : | コロソ |
| | 空白 |

3.2.6 phssyndicrm (同義語辞書削除ユティリティ)

(1) 機能

phssyndicreg で登録した同義語辞書を削除します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phssyndicrm 同義語辞書名 [-u {jis | ms | jis_ucs4 | ms_ucs4}]
```

(4) コマンド引数

同義語辞書名

削除する同義語辞書の同義語辞書名を指定します。登録されている名称を指定してください。

-u

削除する Unicode 種別を指定します。データベースの文字コード種別が utf-8 の場合は、必ず指定します。データベースの文字コードが utf-8 でない場合には指定しないでください。

次のどれかを指定してください。

- jis : JISX0221 で UCS2 の範囲
- ms : MS-Unicode で UCS2 の範囲
- jis_ucs4 : JISX0221 で UCS4 の範囲
- ms_ucs4 : MS-Unicode で UCS4 の範囲

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

SGML プラグインを登録したすべての HiRDB 運用環境から、同一の同義語辞書名の同義語辞書を削除してください。

3.2.7 phssyndicconv (同義語辞書移行ユティリティ)

(1) 機能

同義語辞書を Text Search Plug-in 02-01 以降のバージョン形式に移行します。また、次に示す条件をすべて満たす場合は、同義語辞書移行ユティリティによる同義語辞書の移行が必要です。

- Text Search Plug-in 02-01 よりも前のバージョンで作成した同義語辞書を使用する場合
- 大小文字・全角文字半角文字統一機能または拗音統一機能、および濁音統一機能を利用する場合

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phssyndicconv 同義語辞書名

(4) コマンド引数

同義語辞書名

移行する同義語辞書の同義語辞書名を指定します。登録されている名称を指定します。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

同義語辞書を移行する場合、SGML プラグインを登録したすべての HiRDB 運用環境に対して、実施してください。

3.2.8 phssyndicbkup (同義語辞書バックアップユティリティ)(UNIXの場合)

(1) 機能

登録した同義語辞書のバックアップを取得します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phssyndicbkup バックアップファイル名またはデバイス名

(4) コマンド引数

バックアップファイル名またはデバイス名

バックアップファイル名またはデバイス名を指定します。

(5) 規則

このユティリティは、必ず HiRDB を停止してから実行してください。

(6) 注意事項

02-01 よりも前のバージョンで作成した同義語辞書をバックアップする場合、次のメッセージが表示される場合がありますが、バックアップ作業には問題ありません。

01-00 ~ 01-04 の場合

```
tar: ディレクトリ名/dic_c/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_w/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_cw/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_y/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_cy/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_wy/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_cwy/* の状態がわかりません。 ダンプされません。
```

01-05 ~ 02-00 の場合

```
tar: ディレクトリ名/dic_y/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_cy/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_wy/* の状態がわかりません。 ダンプされません。  
tar: ディレクトリ名/dic_cwy/* の状態がわかりません。 ダンプされません。
```

ディレクトリ名: \$PDDIR/plugin/_phsgml

同義語辞書のリストアには、HiRDB 管理者が、次のコマンドを実行してください。

```
tar xvf バックアップファイル名またはデバイス名
```

リストアするとき、すでに同じ名称の同義語辞書がある場合は、登録済みの同義語辞書が上書きされます。

3.2.9 phnidxls (インデクス情報の取得ユーティリティ)

(1) 機能

指定された n-gram インデクスの次の情報を取得します。

| 情報名 | 内容 |
|--------------------------|--|
| Index Type | インデクス種別。使用するインデクス種別によって、表示内容が異なります。 概念検索用インデクス："CONCEPT" インデクス作成ルール追加機能："ALL" または "MEDIUM" インデクス作成ルール追加機能と概念検索用インデクスを組み合わせた場合："CONCEPT_ALL" または "CONCEPT_MEDIUM" 英文インデクス："ENGLISH" または "ENGLISH_STANDARD" 文字単位インデクス作成機能："GRAM1" |
| Unicode Type | Unicode 種別。文字コード種別が utf-8 の場合だけ表示します。使用する Unicode 種別によって、表示内容が異なります。 UCS2："INDEX_UCS4 指定なし" UCS4："INDEX_UCS4 指定あり" |
| Concept Terms | 特徴ターム取得フラグ。 <ul style="list-style-type: none"> 特徴タームを返却する RD エリアの場合 ON を表示します。 特徴タームを返却しない RD エリアの場合 OFF を表示します。 |
| RD Area Size | RD エリア容量 |
| Index File Using Size | ファイル使用容量 |
| Free Size of Index File | インデクス空き容量 |
| Available Work Area Size | 使用可能ワーク領域サイズ |
| Same Case | 大小文字統一指定。CREATE INDEX のプラグインオプションに「SAMECASE=ON」、 「SAMEWIDE=ON」、 「SAMED=ON」または 「SAMEY=ON」を指定した場合だけ表示します。 |
| Same Wide | 全角文字半角文字統一指定。CREATE INDEX のプラグインオプションに「SAMEWIDE=ON」、 「SAMECASE=ON」、 「SAMED=ON」または 「SAMEY=ON」を指定した場合だけ表示します。 |
| Same Youon | 拗音統一指定。CREATE INDEX のプラグインオプションに「SAMEY=ON」、 「SAMECASE=ON」、 「SAMED=ON」または 「SAMEWIDE=ON」を指定した場合だけ表示します。 |
| Same Dakuon | 濁音統一指定。CREATE INDEX のプラグインオプションに「SAMED=ON」、 「SAMECASE=ON」、 「SAMEY=ON」または 「SAMEWIDE=ON」を指定した場合だけ表示します。 |
| Delete Character | 削除文字指定。 <ul style="list-style-type: none"> 指定あり ON を表示します。 指定なし OFF を表示します。全文インデクスまたは概念インデクスの場合だけ表示します。 |

3. Text Search Plug-in のユティリティ

| 情報名 | 内容 |
|--|---|
| Number of Term | トライノード数 |
| Number of Term for Increment | インクリメント対象トライノード数 |
| Index Size for Increment | 余剰インデクス容量 |
| Specified File Size for Sub Index | CREATE INDEX の実行時にユーザが指定した差分インデクス容量 |
| Sub Index File Using Size | 差分インデクス使用容量 |
| Number of Documents for Index | 登録文書数 |
| Number of Documents for Sub-Index | 差分登録文書数 |
| Ratio of Condensable Documents | 詰め替え可能文書率 |
| Ratio of Condensed Index | 詰め替え作業率 |
| Number of No Condensed Index | 詰め替え未実施インデクス数 |
| Delay Status | インデクス分割遅延作成状態 |
| Size of Delay File (Using ratio) | インデクス分割遅延作成用ファイル容量および使用率 |
| Number of Documents for Unfinished-Index | インデクス未登録文書数。CREATE INDEX 実行時のプラグインオプションに初期分割登録，または分割遅延登録を定義しない場合には，常に 0 を表示します。 |
| Size of Original Concept File | 概念検索用統計情報サイズ |
| Max Page of Original Concept File | 概念検索用統計情報ファイル容量 |
| Size of Integration Concept File | 統合概念検索用統計情報サイズ。「phnconrstr (統計情報反映ユティリティ)」を実行した場合に表示します。 |
| Max Page of Integration Concept File | 統合概念検索用統計情報ファイル容量。「phnconrstr (統合統計情報反映ユティリティ)」を実行した場合に表示します。 |

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phnidxls -d インデクス識別子 -r RDエリア名 [-w 同時実行指定] [-q 世代番号]

(4) コマンド引数

-d インデクス識別子

インデクス情報取得対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスが存在する RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-w 同時実行指定

インデクス情報の取得ユティリティと検索処理を同時に実行するかどうかを指定します。省略時は、"wait" が設定されます。

- wait : 同時実行しません。
- nowait : 同時実行します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

3. Text Search Plug-in のユティリティ

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合に有効です。サーバ側定義が同時実行となっていない場合は、同時実行にはなりません。

3.2.10 phnincmt (インデクスのインクリメンタルユティリティ)

(1) 機能

n-gram インデクスをインクリメンタル (最適化) します。インクリメンタルは、容量が大きいため検索に時間が掛かるインデクスに対して、そのインデクスのキー (n-gram) の文字数を増やし、容量の小さなインデクスを作成することで検索性能を向上させる処理です。

インクリメンタル処理は、RD エリア名単位で指定します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phnincmt -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-t インクリメンタル実行時間] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インクリメンタル対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスが存在する RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

3. Text Search Plug-in のユティリティ

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

インクリメンタル処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インクリメンタル処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インクリメンタル処理前およびインクリメンタル処理後にバックアップを取得する必要があります。

-t インクリメンタル実行時間

インデクスのインクリメンタルユティリティの実行時間を分単位で指定します。実行時間は半角数字で 0 ~ 6,000 の間で 15 分単位で指定してください。15 の倍数でない値が指定された場合は、15 分単位で切り上げられます。実行時間に 0 を指定した場合、またはこのオペランドを省略した場合、インクリメンタル処理が完了するまで処理を続けます。指定した実行時間でインクリメンタル処理が完了しなかった場合、ユティリティが指定時間を過ぎて終了した旨のメッセージを出力します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

インクリメンタル実行時間を指定して、インデクスのインクリメンタルユティリティを実行した場合、インクリメント対象トライノード数、および余剰インデクス容量が 0 になるまでこのユティリティを実行してください。

インクリメント対象トライノード数、および余剰インデクス容量は、「phnidxls (インデクス情報の取得ユティリティ)」で確認してください。

インデクス未登録状態の文書は、インクリメンタルの対象外となります。

インクリメンタルユティリティ実行後、「phnstartidx (インデクス作成開始ユティリティ)」を実行した場合、必要に応じて文字数の多いインデクスを自動的に作成します。

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

インクリメンタル実行時間を指定した場合、指定時間と実際の処理時間との間に多少のずれが生じることがあります。

インクリメンタル実行時間を指定した場合、指定時間を過ぎてインクリメンタル処理を中断して終了した場合は、メッセージログに「02001-I phnincmt is Time Out.」のメッセージを出力します。インクリメンタル処理を完了して終了した場合は、「02003-I phnincmt command ended.」のメッセージを出力します。

このユティリティは、英文検索機能を使用することはできません。

UCS4 インデクスの場合は、処理を行わずに正常終了します。

3.2.11 phnmerge (インデクスのマージユティリティ)

(1) 機能

差分インデクスに登録されているインデクスを、MASTER インデクスにマージします。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phnmerge -d インデクス識別子 [-l ログ取得方式] [-w 同時実行指定] [-q 世代番号]

(4) コマンド引数

-d インデクス識別子

マージ処理対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

マージ処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。マージ処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。マージ処理前およびマージ処理後に

バックアップを取得する必要があります。

-w 同時実行指定

インデクスのマージユティリティと検索処理を同時に実行するかどうかを指定します。省略時は、"wait" が設定されます。

- wait : 同時実行しません。
- nowait : 同時実行します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合に有効です。サーバ側定義が同時実行となっていない場合は、同時実行とはなりません。

UCS4 インデクスの場合は、処理を行わずに正常終了します。

3.2.12 phnexpnd (インデクスの拡張ユティリティ)

(1) 機能

インデクスファイルを RD 空きエリアに応じて作成し、インデクス全体のサイズを拡張します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phnexpnd -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-q 世代番号]

(4) コマンド引数

-d インデクス識別子

拡張対象のインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスがある RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

インデクス拡張処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。

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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。

ロールバックに必要なデータベース更新ログを取得します。インデクス拡張処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。

データベース更新ログを取得しません。インデクス拡張処理前およびインデクス拡張処理後にバックアップを取得する必要があります。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

3.2.13 phnstartidx (インデクス作成開始ユティリティ)

(1) 機能

インデクス作成方法に初期分割登録, または分割遅延登録を指定して作成したインデクスに対して, インデクス未登録状態の文書をインデクスに登録します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phnstartidx -d インデクス識別子 [-t インデクス作成実行時間] [-l ログ取得方式] [-c 登録件数] [-w 同時実行指定] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インデクス登録対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は, 英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。なお, sh, csh および ksh の場合は, さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は, 英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。

-t インデクス作成実行時間

インデクス作成開始ユティリティの実行時間を分単位で指定します。

指定範囲は, 0 ~ 6,000 となります。0 を指定した場合, またはこのオペランドの指定を省略した場合は, 時間では制限されません。

指定を省略した場合は, 実行の可否を問い合わせるメッセージを表示します。なお, 分割遅延登録・更新系 SQL 同時実行機能を利用する場合, このオプションは指定できません。

-t 指定によるインデクス作成実行時間指定は UCS4 インデクス (INDEX_UCS4 指定) の場合は使用できません。

-l ログ取得方式

インデクス登録処理でのログ取得方式を指定します。省略時は, "p" が設定されます。

分割遅延登録・更新系 SQL 同時実行機能を利用する場合は, "a" または "p" を指定してください。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデクス登録処理後, バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデクス登録処理前およびインデクス登録処理後にバックアップを取得する必要があります。

-c 登録件数

インデクス作成をする件数を, 1 ~ 2,147,483,646 の範囲で指定します。
指定を省略した場合は, 登録件数では制限されません。

-w 同時実行指定

このユティリティと検索処理を同時に実行するかどうか, またはこのユティリティと遅延登録を同時に実行するかどうかを指定します。省略時は, "wait" が設定されます。

- wait : 同時実行しません。
- nowait : 登録・検索を同時実行します。
- nowait_with_update : 登録・検索同時実行に加えて, 分割遅延登録・更新系 SQL を同時実行します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合, 処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は, カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは, Staticizer Option がインストールされていない場合は指定できません。

なお, インナレプリカ機能については, マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは, HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

インデクス未登録状態の文書が存在しないインデクスに対して、このユティリティを実行しても、無効となり正常終了します。

インデクス作成実行時間を指定してインデクス作成開始ユティリティを実行した場合、インデクス未登録状態の文書数が 0 になるまでこのユティリティを実行してください。インデクス未登録文書数は、「phnidxls (インデクス情報の取得ユティリティ)」で確認してください。

-t オプション、-c オプションの指定を省略した場合は、実行の可否を確認するメッセージを表示します。

インデクス作成開始ユティリティは、データの整合性を確保するため、インデクス作成実行時間に指定した時間が経過しても、終了しない場合があります。この場合は、データの整合性を確保できた時点で終了します。インデクス作成実行時間が経過してからユティリティが終了するまでの時間は、n-gram インデクスプラグインの環境定義ファイルに指定した一括登録時最大メモリ所要量とディスク性能に依存します。

インデクス作成実行時間を指定した場合、指定時間を過ぎてインデクス作成処理を中断して終了した場合は、メッセージログに「02001-I phnstartidx is Time Out.」のメッセージを出力します。インデクス作成処理を完了して終了した場合は、「02003-I phnstartidx command ended.」のメッセージを出力します。

-t オプションのインデクス実行時間と -c オプションの登録件数を指定した場合、どちらかの条件が成立した時点で登録処理を終了します。

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合に有効です。サーバ側定義が同時実行となっていない場合は、同時実行とはなりません。

分割した表のインデクスに対し、-c オプションを指定してインデクスを作成する場合、指定した文書数のインデクスを分割した表ごとに作成します。

インデクス作成開始ユティリティは、登録要求件数分の行に対して参照できる共用排他を実行します。このため、分割遅延登録・更新系 SQL 同時実行を指定してこのユティリティを実行する場合は、排他ができるように排他資源を確保してください。

-w オプションに nowait_with_update を指定し、さらに -t オプションを指定した場合、メッセージログに「62014-E Option argument invalid. option=xx」のメッセージを出力します。

使用している HiRDB のバージョンが、この製品が前提としている HiRDB のバージョンと不一致の場合は、エラーメッセージ「KFPL28000-E Internal error occurred」が出力され、異常終了します。

3.2.14 phndelay (インデクス分割遅延登録運用変更ユーティリティ)

(1) 機能

インデクス作成方法に分割遅延登録を指定して作成したインデクスに対して、分割遅延登録運用開始および終了を指定します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phndelay -d インデクス識別子 -m { on|off} [-l ログ取得方式] [-f] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

分割遅延登録運用方法を変更するインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-m

- on : インデクス分割遅延登録運用を開始します。
- off : インデクス分割遅延登録運用を終了します。

-l ログ取得方式

分割遅延登録運用方法変更処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。

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します。インデクスのバックアップを取得する必要はありません。

p :

3. Text Search Plug-in のユティリティ

更新前ログ取得モードで実行します。

ロールバックに必要なデータベース更新ログを取得します。分割遅延登録運用方法変更処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。

データベース更新ログを取得しません。分割遅延登録運用方法変更処理前および分割遅延登録運用方法変更処理後にバックアップを取得する必要があります。

-f

応答メッセージを出力しないで実行する場合に指定します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

分割遅延登録運用の終了が指定された場合、すべての未反映文書を登録します。その際、時間 / 件数による中断はできないので、分割遅延登録運用の終了以前に「phnstartidx (インデクスの作成開始ユティリティ)」を実行して登録処理を完了させておいてください。また、分割遅延登録運用の終了実行時には、応答のメッセージを出力します。-f オプションが指定された場合は、応答のメッセージを出力しません。

分割遅延登録用でないインデクスに対して、このユティリティを実行しても無効となります (エラーになりません。正常終了します)。

分割遅延登録・更新系 SQL 同時実行を使用している場合は変更できません。

3.2.15 phnmodidx (インデクス情報変更ユティリティ)

(1) 機能

作成済みのインデクスに対して定義を変更します。インデクス分割遅延作成ができるインデクスへ移行できます。また、インデクス分割遅延作成用ファイルサイズを拡張できます。

また UCS2 用インデクスから UCS4 用インデクスへの移行を行うことができます。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phnmodidx -d インデクス識別子 -v 変更項目 [-l ログ取得方式] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インデクス情報変更の対象となる、インデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-v 変更項目

変更するインデクス定義項目を、次の形式で指定します。

- "DELAY_SIZE= インデクス分割遅延登録用領域サイズ"

サイズは、2,608 ~ 129,584 キロバイトの範囲で指定してください。新規に移行する際には、指定されたサイズでファイルを作成します。すでにファイルがある場合は、指定されたサイズまでファイルを拡張します。最小拡張単位は 2,048 キロバイトです。端数は切り捨てられます。また、拡張回数の上限は 62 回です。

- "INDEX_TYPE=UCS4"

UCS2 用インデクスから、UCS4 用インデクスへの移行を行いません。

-l ログ取得方式

3. Text Search Plug-in のユティリティ

インデクス情報変更処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデクス情報変更処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデクス情報変更処理前およびインデクス情報変更処理後にバックアップを取得する必要があります。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

このユティリティは、インデクス分割遅延作成用ファイルを作成または拡張します。そのまま実行すると、RD エリア不足が発生することがあります。その場合、作成または拡張するサイズ分の RD エリアを、あらかじめ追加することをお勧めします。

このユティリティで UCS4 用インデクスへの移行を行なう場合、RD エリアの空き容量 (phnidxls コマンドの Available Work Area Size) が十分に存在していることを確認してください。そのまま実行すると、RD エリア不足が発生することがあります。

次のデータベースに対して UCS4 用インデクスへの移行は実行できません。

- SUB_INDEX (差分インデクス)
- ENGLISH, ENGLISH_STANDARD (英文インデクス)
- 文字コード種別が utf-8 以外のデータベース

UCS4 インデクスへの移行時に、既にインデクス作成済み文書の UCS2 の範囲を超える UCS4 の範囲の文字は検索対象となりません。検索対象とするためには、プラグインオプションに INDEX_UCS4 を指定して、インデクスを再作成してください。

このユティリティで分割遅延作成できるインデクスへ移行しても、分割遅延登録運用の開始 / 終了状態は変更されません。運用方法を変更する場合は別途 phndelay コマンドを実行する必要があります。また、現在の運用方法は phnidxls コマンド実行結果の「Delay Status」で確認できます

分割遅延登録・更新系 SQL 同時実行を使用している場合は変更できません。

インデクスの再初期化を行った場合、このユティリティでの移行情報は無効になり、初めにインデクスを定義した状態に戻ります。

3.2.16 phncond (インデクス詰め替えユティリティ)

(1) 機能

不要なインデクスを削除します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phncond -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-t 詰め替え時間] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

詰め替え対象となるインデクス識別子を指定します。

必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。

また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。

¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスがある RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

インデクス詰め替え処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデクス詰め替え処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデクス詰め替え処理前およびインデクス詰め替え処理後にバックアップを取得する必要があります。

-t 詰め替え時間

インデクス詰め替えユティリティの実行時間を分単位で指定します。実行時間は半角数字で 0 ~ 6,000 の間で 15 分単位で指定してください。15 の倍数でない値が指定された場合は、15 分単位で切り上げられます。実行時間に 0 を指定した場合、またはこのオペランドを省略した場合、詰め替え処理が完了するまで処理を続けます。指定した実行時間で詰め替え処理が完了しなかった場合、ユティリティが指定時間を過ぎて終了した旨のメッセージを出力します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

詰め替え処理は、1 回の処理ですべてのインデクスを処理できないことがあります。

3. Text Search Plug-in のユティリティ

このため、詰め替え実施後、「phnidxls (インデクス情報の取得ユティリティ)」で未実施インデクス数 (Number of No Condensed Index) を確認してください。この値が大ききときは、インデクス空き容量に注意して、再度詰め替えを実行してください。

このユティリティは、インデクス空き容量がある状態 (推奨 10% 以上) で実行する必要があります。空きのない状態でユティリティを実行したとき、または詰め替え未実施インデクス数 (Number of No Condensed Index) が減らない場合は、新たに RD エリアを追加して「phnexpnd (インデクスの拡張ユティリティ)」を実行し、空き容量がある状態で実行してください。空き容量は、「phnidxls (インデクス情報の取得ユティリティ)」の Free Size of Index File で確認してください。

詰め替え実行時間を指定した場合、指定時間と実際の処理時間との間に多少のずれが生じることがあります。

文書を削除すると、詰め替え可能文書率は増加します。このため、詰め替え中断時に文書を削除すると、詰め替え処理完了後でも詰め替え可能文書率は 0 になりません。

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

差分インデクスは詰め替え処理の対象となりません。

詰め替えを実行する契機の指標として、「phnidxls (インデクス情報の取得ユティリティ)」の「Ratio of Condensable Documents (詰め替え可能文書率)」が、50% 以上を目安としてください。

「phnidxls (インデクス情報の取得ユティリティ)」の「Ratio of Condensed Index (詰め替え作業率)」は、詰め替え処理が制限時間で終了したときの詰め替え作業進捗率を表します。

このユティリティを実行する際に、グローバルバッファを指定すると、実行性能が向上します。

詰め替え時間を指定した場合、指定時間を過ぎて詰め替え処理を中断したときは、メッセージログに「02001-I phncond is Time Out.」のメッセージが出力されます。詰め替え処理を完了して終了した場合は、「02003-I phncond command ended」のメッセージが出力されます。

3.2.17 phnconget (統計情報収集ユーティリティ)

(1) 機能

概念検索用統計情報を収集し、指定したディレクトリに出力します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

`phnconget -d インデクス識別子 [-r RDエリア名] -o 統計情報出力先定義ファイル名 [-w 同時実行指定] -m 収集モード指定 [-q 世代番号]`

(4) コマンド引数

-d インデクス識別子

統計情報収集の対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定したインデクスがある RD エリア名を指定します。省略した場合は、対象となるインデクスを構成するすべての RD エリアの統計情報を出力します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

3. Text Search Plug-in のユティリティ

-o 統計情報出力先定義ファイル名

統計情報の出力先を定義したファイル名を絶対パスで指定します。

-w 同時実行指定

統計情報収集ユティリティと検索処理を同時に実行するかどうかを指定します。省略時は、"wait" が設定されます。

- wait : 同時実行しません。
- nowait : 同時実行します。

-m 収集モード指定

統計情報の収集方法を指定します。

- normal : 概念検索インデクス中の統計情報をそのまま収集します。
- all : インデクス作成ルール追加機能を使用した場合と同等の統計情報を収集します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 統計情報出力先定義ファイル

統計情報出力先定義ファイルには、収集した統計情報の出力先を BES 単位に指定します。また、指定した以外の BES からの統計情報の出力先として、デフォルトの出力先を指定します。BES 単位の出力先とデフォルトの出力先のどちらの指定もない場合は、ユティリティ実行時にエラーとなります。

統計情報の出力先として指定するディレクトリは、ユティリティ実行前に処理対象となる各 BES マシンに作成しておく必要があります。出力先として指定したディレクトリが存在しない場合は、エラーとなり、処理が中断されます。指定した出力先に、すでに同じ名称の統計情報出力ファイルが存在している場合は、新しい統計情報で上書きされません。

(a) 記述形式

BES 単位に出力先を指定する場合

[SERVERS]

サーバ名称 = 出力先ディレクトリ絶対パス名

デフォルトの出力先を指定する場合

[DEFAULT]

DEFAULT= 出力先ディレクトリ絶対パス名

(b) 定義例

統計情報出力先定義ファイルの定義例を図 3-1 および図 3-2 に示します。

図 3-1 統計情報出力先定義ファイルの定義例 (UNIX の場合)

```
[SERVERS]
bes01=/home/HiRDB_TSPlugin/ConDir01
bes02=/home/HiRDB_TSPlugin/ConDir02
bes03=/home/HiRDB_TSPlugin/ConDir03
      :
      :
      :
[DEFAULT]
DEFAULT =/home/HiRDB_TSPlugin/ConDir
```

図 3-2 統計情報出力先定義ファイルの定義例 (Windows の場合)

```
[SERVERS]
bes01=C:\work\HiRDB_TSPlugin\ConDir01
bes02=C:\work\HiRDB_TSPlugin\ConDir02
bes03=C:\work\HiRDB_TSPlugin\ConDir03
      :
      :
      :
[DEFAULT]
DEFAULT =C:\work\HiRDB_TSPlugin\ConDir
```

(6) 規則

このユーティリティは、HiRDB が稼働中のときだけ実行できます。

(7) 注意事項

このユーティリティは、分割されている概念検索用の統計情報を統合する処理の一環として、各分割インデクス中のおおのこの統計情報を、指定したディレクトリに出力します。最終的には、一つのインデクスを構成するすべての統計情報を出力して、「phnconbuild (統計情報統合ユーティリティ)」で統合する必要があります。

インデクス作成ルール追加機能を使用しないで作成したインデクスに対して、収集モードに "all" を指定した場合は、登録文書から統計情報を作成します。このため、収集モード指定で "normal" を指定した場合に比べて、処理に時間が掛かります。

統計情報出力先定義ファイルに指定したディレクトリ下には、「"NGRAMTRIE"+ 通番」という名称でファイルが出力されます。

ユーティリティ実行時にエラーが発生した場合、作成中の統計情報が残る場合があります。

3. Text Search Plug-in のユティリティ

す。その場合、統計情報出力先定義ファイルに指定したディレクトリ内のファイルをすべて削除してください。

遅延更新や差分インデックスを使用している場合は、インデックスへ反映してから実行してください。未反映文書の統計情報が作成されないことがあります。

有効文書が 0 件のインデックスに対して、収集モード指定 = "all" を指定した場合、統計情報は作成されません。

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合に有効です。サーバ側定義が同時実行となっていない場合は、同時実行とはなりません。

英文検索機能のユティリティは指定できません。

文字コード種別が utf-8 の場合、このユティリティは指定できません。

3.2.18 phnconbuild (統計情報統合ユティリティ)

(1) 機能

概念検索用統計情報を統合します。

「phnconget (統計情報収集ユティリティ)」で収集した各統計情報ファイルを統合し、統合した統計情報ファイルを出力します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phnconbuild -i 入力統計情報ファイル格納先定義ファイル名 -o 統合統計情報出力先ファイル名

(4) コマンド引数

-i 入力統計情報ファイル格納先定義ファイル名

入力する統計情報ファイルの格納先を定義したファイル名を絶対パスで指定します。
この定義ファイルに指定したすべての統計情報ファイルの内容を統合します。

-o 統合統計情報出力先ファイル名

統合した統計情報を出力するファイル名を絶対パスで指定します。

(5) 入力統計情報ファイル格納先定義ファイル

入力統計情報ファイル格納先定義ファイルには、入力する統計情報ファイルの格納先をファイル単位に指定します。デフォルトの格納先も指定できます。

入力統計情報ファイルとして同じファイルを複数回指定した場合は、異なる統計情報として処理されます。また、入力する統計情報ファイルの数が0の場合はエラーとなりません。

(a) 記述形式

統計情報ファイル単位に格納先を指定する場合
[FILE]

FILE= 統計情報ファイル絶対パス名

デフォルトの格納先を指定する場合

[DEFAULT]

DEFAULT= 統計情報ファイル格納ディレクトリ絶対パス名

(b) 定義例

入力統計情報ファイル格納先定義ファイルの定義例を図 3-3 および図 3-4 に示します。

3. Text Search Plug-in のユティリティ

図 3-3 入力統計情報ファイル格納先定義ファイルの定義例 (UNIX の場合)

```
[FILE]
FILE=/home/HiRDB_TSPugin/ConDir01/NGRAMTRIE1
FILE=/home/HiRDB_TSPugin/ConDir02/NGRAMTRIE2
FILE=/home/HiRDB_TSPugin/ConDir03/NGRAMTRIE3
      :
      :
[DEFAULT]
DEFAULT =/home/HiRDB_TSPugin/ConDir/
```

図 3-4 入力統計情報ファイル格納先定義ファイルの定義例 (Windows の場合)

```
[FILE]
FILE=C:¥work¥HiRDB_TSPugin¥ConDir01¥NGRAMTRIE1
FILE=C:¥work¥HiRDB_TSPugin¥ConDir02¥NGRAMTRIE2
FILE=C:¥work¥HiRDB_TSPugin¥ConDir03¥NGRAMTRIE3
      :
      :
[DEFAULT]
DEFAULT =C:¥work¥HiRDB_TSPugin¥ConDir
```

(6) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(7) 注意事項

このユティリティを実行する際には、インデクス一つ分の統計情報を入力としてください。

一つのインデクスを構成するすべての分割インデクスが入力として指定されたかどうかの過不足はチェックしません。異なるインデクスの統計情報を指定した場合は、同じインデクスの情報として統合します。

英文検索機能のユティリティは指定できません。

文字コード種別が utf-8 の場合、このユティリティは指定できません。

3.2.19 phnconrstr (統計情報反映ユティリティ)

(1) 機能

統合した概念検索用統計情報をデータベースに反映します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phnconrstr -d インデクス識別子 [-r RDエリア名] -i 統計情報定義ファイル名
[-l ログ取得方式] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

統合統計情報反映の対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定したインデクスがある RD エリア名を指定します。省略した場合は、対象となるインデクスを構成するすべての RD エリアの統計情報を更新します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

3. Text Search Plug-in のユティリティ

-i 統計情報定義ファイル名

統合した統計情報ファイル名を定義したファイル名を絶対パスで指定します。この定義ファイルに指定した統計情報ファイルの内容をデータベースに反映します。

-l ログ取得方式

統計情報反映でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。統計情報反映後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。データベース更新ログを取得しません。
統計情報反映後にバックアップを取得する必要があります。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 統計情報定義ファイル

統計情報定義ファイルには、データベースへ反映する統合統計情報のファイル名を BES 単位に指定します。また、デフォルトの統合統計情報ファイルを指定できます。統合統計情報のファイル名を指定しなかった場合には、デフォルトの統合統計情報ファイルをデータベースへ反映します。BES 単位の反映元とデフォルトの反映元のどちらの指定もない場合は、ユティリティ実行時にエラーとなります。

(a) 記述形式

BES 単位に反映先を指定する場合

[SERVERS]

サーバ名称 = 統合統計情報ファイル絶対パス名

デフォルトのファイルを指定する場合

[DEFAULT]

DEFAULT= 統合統計情報ファイル絶対パス名

(b) 定義例

統計情報定義ファイルの定義例を図 3-5 および図 3-6 に示します。

図 3-5 統計情報定義ファイルの定義例 (UNIX の場合)

```
[SERVERS]
bes01=/home/HiRDB_TSPlugin/conmergefile
bes02=/home/HiRDB_TSPlugin/conmergefile
bes03=/home/HiRDB_TSPlugin/conmergefile
      :
      :
      :
[DEFAULT]
DEFAULT =/home/HiRDB_TSPlugin/conmergefile
```

図 3-6 統計情報定義ファイルの定義例 (Windows の場合)

```
[SERVERS]
bes01=C:¥work¥HiRDB_TSPlugin¥conmergefile
bes02=C:¥work¥HiRDB_TSPlugin¥conmergefile
bes03=C:¥work¥HiRDB_TSPlugin¥conmergefile
      :
      :
      :
[DEFAULT]
DEFAULT =C:¥work¥HiRDB_TSPlugin¥conmergefile
```

(6) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(7) 注意事項

このユティリティ実行前に、「phnconbuild (統計情報統合ユティリティ)」で統合した統計情報ファイルを各 BES マシンに配布してください。

統計情報の反映時には、統合した統計情報を格納する領域を RD エリア内に作成します。そのため、ユティリティ実行前に、RD エリアに空き領域があるかどうかを確認する必要があります。空き領域の確認には HiRDB の運用コマンド「pddbls (RD エリアの状態表示)」を使用します。

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

該当インデクスへのユティリティ実行が 2 回目以降の場合には、作成済みの統計情報

3. Text Search Plug-in のユティリティ

格納領域に上書きして反映します。また、格納する統計情報のサイズが前回実行時より増加している場合は、4,188 ページ単位で統合概念検索用統計情報用ファイル容量を拡張します。RD エリアの残容量に注意してください。

統計情報の反映は、RD エリア単位で実行することをお勧めします。

英文検索機能のユティリティは指定できません。

文字コード種別が utf-8 の場合、このユティリティは指定できません。

3.2.20 phnchgcon (特徴ターム取得 RD エリア指定ユティリティ)

(1) 機能

横分割した表に対して、特徴ターム抽出関数を使用すると、インデクスを構成するすべての RD エリアから特徴タームを抽出します。

このユティリティは、特徴タームを取得する RD エリアを一つに限定します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

特徴ターム取得 RD エリアを指定する場合

```
phnchgcon -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-q 世代番号]
```

特徴ターム取得 RD エリアの指定を解除する場合

```
phnchgcon -d インデクス識別子 -m TermGet [-l ログ取得方式] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

特徴ターム取得 RD エリアを指定、または指定を解除するインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

特徴タームの取得対象とする RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。ま

3. Text Search Plug-in のユティリティ

た、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

特徴ターム取得 RD エリアの指定 / 解除でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。特徴ターム取得 RD エリアの指定 / 解除後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。特徴ターム取得 RD エリアの指定 / 解除後にバックアップを取得する必要があります。

-m TermGet

特徴ターム取得 RD エリアの指定を解除する場合に指定します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

特徴タームを取得する RD エリアを変更する場合は「phnchgcon (特徴ターム取得 RD エリア指定ユティリティ)」を再度実行してください。

英文検索機能のユティリティは指定できません。

3.3 文字列検索プラグインのユティリティ

この節では、Text Search Plug-in の文字列検索プラグイン (FREEWORD プラグインおよび IXFREEWORD プラグイン) で、文字列データを扱うために必要なユティリティの詳細について説明します。

3.3.1 phcsyndicreg (同義語辞書登録ユティリティ)

(1) 機能

同義語検索をするために必要な同義語辞書を登録します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phcsyndicreg 同義語辞書名 同義語辞書テキストファイル名 [-u {ms |ms_ucs4}]

(4) コマンド引数

同義語辞書名

辞書を登録する際に付ける名称を指定します。システム内で区別できるユニークな名称を、1 ~ 32 バイトの範囲内で指定してください。

同義語辞書テキストファイル名

同義語辞書のテキストファイルの名称を指定します。
ユティリティを実行するときのカレントディレクトリにあるファイル名または絶対パスを含めたファイル名を指定します。

-u

Unicode 種別を指定します。データベースの文字コード種別が utf-8 の場合は、必ず指定します。データベースの文字コードが utf-8 でない場合には指定しないでください。次のどちらかを指定してください。

- ms : MS-Unicode で UCS2 の範囲
- ms_ucs4 : MS-Unicode で UCS4 の範囲

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

FREEWORD プラグインを登録したすべての HiRDB 運用環境に対して、同一の同義語辞書テキストファイルを使用し、同一の同義語辞書名で登録してください。

文書検索プラグインと同じ同義語辞書を使用する場合は、文書検索プラグインで登録

した同義語辞書テキストファイルを、文字列検索プラグインにも同様に登録してください。

同義語辞書名に指定できる文字は、HiRDB をセットアップしたときの文字コード、記号英数字、および半角カタカナです。なお、同義語辞書名に指定できない記号および英数字については、「3.2.5 phssyndicreg (同義語辞書登録ユティリティ)」の表 3-2 を参照してください。

データベースの文字コード種別が utf-8 の場合、UCS4 の範囲のインデクス (INDEX_UCS4) に対する同義語検索は、UCS4 の範囲で作成した同義語辞書を使用します。INDEX_UCS4 指定のないインデクスは、UCS2 の範囲で作成した同義語辞書を使用します。また、extract 関数は UCS4 の範囲の同義語辞書を優先的に使用しますが、UCS4 の範囲の同義語辞書が作成されていない場合には、UCS2 の範囲の同義語辞書を使用します。

系切り替え運用を行なう場合は、「付録 E 系切り替え」を参照してください。

3.3.2 phcsyndicrm (同義語辞書削除ユティリティ)

(1) 機能

phcsyndicreg で登録した同義語辞書を削除します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phcsyndicrm 同義語辞書名 [-u {ms |ms_ucs4}]
```

(4) コマンド引数

同義語辞書名

削除する同義語辞書の同義語辞書名を指定します。登録されている名称を指定してください。

-u

削除する Unicode 種別を指定します。データベースの文字コード種別が utf-8 の場合は、必ず指定します。データベースの文字コードが utf-8 でない場合には指定しないでください。

次のどちらかを指定してください。

- ms : MS-Unicode で UCS2 の範囲
- ms_ucs4 : MS-Unicode で UCS4 の範囲

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

FREWORD プラグインを登録したすべての HiRDB 運用環境から、同一の同義語辞書名の同義語辞書を削除してください。

3.3.3 phcsyndicconv (同義語辞書移行ユティリティ)

(1) 機能

同義語辞書を Text Search Plug-in 02-01 以降のバージョン形式に移行します。次の条件をすべて満たす場合、同義語辞書移行ユティリティによる同義語辞書の移行が必要となります。

- Text Search Plug-in 02-01 よりも前のバージョンで作成した同義語辞書を使用する場合
- 大小文字・全角文字半角文字統一機能または拗音統一機能、および濁音統一機能を利用する場合

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phcsyndicconv 同義語辞書名

(4) コマンド引数

同義語辞書名

移行する同義語辞書の同義語辞書名を指定します。登録されている名称を指定してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

同義語辞書を移行する場合には、FREEWORD プラグインを登録したすべての HiRDB 運用環境に対して、実施してください。

3.3.4 phcsyndicbkup (同義語辞書バックアップユティリティ)(UNIXの場合)

(1) 機能

登録した同義語辞書のバックアップを取得します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phcsyndicbkup バックアップファイル名またはデバイス名

(4) コマンド引数

バックアップファイル名またはデバイス名

バックアップファイル名またはデバイス名を指定します。

(5) 規則

このユティリティは、必ず HiRDB を停止してから実行してください。

(6) 注意事項

02-01 よりも前のバージョンで作成した同義語辞書をバックアップする場合、次のメッセージが表示される場合がありますが、バックアップ作業には問題ありません。

01-00 ~ 01-04 の場合

```
tar: ディレクトリ名/dic_c/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_w/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_cw/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_y/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_cy/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_wy/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_cwy/* の状態がわかりません。ダンプされません。
```

01-05 ~ 02-00 の場合

```
tar: ディレクトリ名/dic_y/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_cy/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_wy/* の状態がわかりません。ダンプされません。  
tar: ディレクトリ名/dic_cwy/* の状態がわかりません。ダンプされません。
```

ディレクトリ名: \$PDDIR/plugin/_phcfwd

同義語辞書のリストアには、HiRDB 管理者が、次のコマンドを実行してください。

```
tar xvf バックアップファイル名またはデバイス名
```

リストアする時に、すでに同じ名称の同義語辞書がある場合は、登録済みの同義語辞書が上書きされます。

3.3.5 phidxls (インデクス情報の取得ユティリティ)

(1) 機能

指定された IXFREEWORD インデクスの次の情報を取得します。

| 情報 | 内容 |
|--------------------------------|---|
| Index Type | 英文インデクス: "ENGLISH" または "ENGLISH_STANDARD" 文字単位インデクス作成機能: "GRAM1" |
| Unicode Type | Unicode 種別。文字コード種別が UTF8 の場合だけ表示します。使用する Unicode 種別によって、表示内容が異なります。 UCS2: "INDEX_UCS4 指定なし" UCS4: "INDEX_UCS4 指定あり" |
| RD Area Size | RD エリア容量 |
| Index File Using Size | ファイル使用容量 |
| Free Size of Index File | インデクス空き容量 |
| Available Work Area Size | 使用可能ワーク領域サイズ |
| Same Case | 大小文字統一指定。CREATE INDEX のプラグインオプションに「SAMECASE=ON」,「SAMEWIDE=ON」,「SAMED=ON」または「SAMEY=ON」を指定した場合だけ表示します。 |
| Same Wide | 全角文字半角文字統一指定。CREATE INDEX のプラグインオプションに「SAMEWIDE=ON」,「SAMECASE=ON」,「SAMED=ON」または「SAMEY=ON」を指定した場合だけ表示します。 |
| Same Youon | 拗音統一指定。CREATE INDEX のプラグインオプションに「SAMEY=ON」,「SAMECASE=ON」,「SAMED=ON」または「SAMEWIDE=ON」を指定した場合だけ表示します。 |
| Same Dakuon | 濁音統一指定。CREATE INDEX のプラグインオプションに「SAMED=ON」,「SAMECASE=ON」,「SAMEY=ON」または「SAMEWIDE=ON」を指定した場合だけ表示します。 |
| Delete Character | 削除文字指定。 <ul style="list-style-type: none"> 指定あり ON を表示します。 指定なし OFF を表示します。全文インデクス, 概念インデクスの場合だけ表示します。 |
| Number of Term | トライノード数 |
| Number of Term for Increment | インクリメント対象トライノード数 |
| Index Size for Increment | 余剰インデクス容量 |
| Number of Documents for Index | 登録文書数 |
| Ratio of Condensable Documents | 詰め替え可能文書率 |

3. Text Search Plug-in のユティリティ

| 情報 | 内容 |
|--|---|
| Ratio of Condensed Index | 詰め替え作業率 |
| Number of No Condensed Index | 詰め替え未実施インデクス数 |
| Delay Status | インデクス分割遅延作成状態 |
| Size of Delay File (Using ratio) | インデクス分割遅延作成用ファイル容量および使用率 |
| Number of Documents for Unfinished-Index | インデクス未登録文書数。CREATE INDEX 実行時のプラグインオプションに初期分割登録, または分割遅延登録を定義しない場合には, 常に 0 を表示します。 |

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

phidxls -d インデクス識別子 -r RDエリア名 [-w 同時実行指定] [-q 世代番号]

(4) コマンド引数

-d インデクス識別子

インデクス情報取得対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は, 英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。なお, sh, csh および ksh の場合は, さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は, 英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。

-r RD エリア名

指定されたインデクスが存在する RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は, 英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また, RD エリア名にスペースが含まれる場合は, " で囲む必要があります。なお, sh, csh および ksh の場合は, さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-w 同時実行指定

インデックス情報の取得ユティリティと検索処理を、同時に実行するかどうかを指定します。省略時は、"wait" が設定されます。

- wait：同時実行しません。
- nowait：同時実行します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0：オリジナル RD エリアを対象とします。
- 1 ~ 10：指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合に有効です。サーバ側定義が同時実行となっていない場合は、同時実行とはなりません。

ユティリティを実行して取得するインデックス種別は、ENGLISH の指定時は ENGLISH、ENGLISH_STANDARD の指定時は ENGLISH_STANDARD と表示します。

3.3.6 phxincmt (インデクスのインクリメンタルユティリティ)

(1) 機能

IXFREEWORD インデクスをインクリメンタル (最適化) します。インクリメンタルは、容量が大きいため検索に時間が掛かるインデクスに対して、そのインデクスのキーの文字数を増やし、容量の小さなインデクスを作成することで検索性能を向上させる処理です。

インクリメンタル処理は、RD エリア名単位で指定します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxincmt -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-t インクリメンタル実行時間] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インクリメンタル対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスが存在する RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

インクリメンタル処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。

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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。

ロールバックに必要なデータベース更新ログを取得します。インクリメンタル処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。

データベース更新ログを取得しません。インクリメンタル処理前およびインクリメンタル処理後にバックアップを取得する必要があります。

-t インクリメンタル実行時間

インデクスのインクリメンタルユティリティの実行時間を分単位で指定します。

実行時間は半角数字で 0 ~ 6,000 の間で 15 分単位で指定してください。15 の倍数でない値が指定された場合は、15 分単位で切り上げられます。

実行時間に 0 を指定した場合、またはこのオペランドを省略した場合、インクリメンタル処理が完了するまで処理を続けます。

指定した実行時間でインクリメンタル処理が完了しなかった場合、ユティリティが指定時間を過ぎて終了した旨のメッセージを出力します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

インクリメンタル実行時間を指定してインデクスのインクリメンタルユティリティを実行した場合、インクリメント対象トライノード数、および余剰インデクス容量が 0 になるまでこのユティリティを実行してください。

インクリメント対象トライノード数、および余剰インデクス容量は、「phxidxls (インデクス情報の取得ユティリティ)」で確認してください。

インデクス未登録状態の文書は、インクリメンタルの対象外となります。

インクリメンタルユティリティ実行後、「phxstartidx (インデクス作成開始ユティリティ)」を実行した場合、必要に応じて文字数の多いインデクスを自動的に作成します。

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

インクリメンタル実行時間を指定した場合、指定時間と実際の処理時間との間に多少のずれが生じることがあります。

インクリメンタル実行時間を指定した場合、指定時間を過ぎてインクリメンタル処理を中断して終了した場合は、メッセージログに「02001-I phxincmt is Time Out.」のメッセージを出力します。インクリメンタル処理を完了して終了した場合は、「02003-I phxincmt command ended.」のメッセージを出力します。

UCS4 のインデクスの場合は、処理を行わずに正常終了します。

3.3.7 phxexpnd (インデクスの拡張ユティリティ)

(1) 機能

インデクスファイルを RD 空きエリアに応じて作成し、インデクス全体のサイズを拡張します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxexpnd -d インデクス識別子 -r RDエリア名 [-l ログ取得方式] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

拡張対象のインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスが存在する RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

3. Text Search Plug-in のユティリティ

-l ログ取得方式

インデックス拡張処理でのログ取得方式を指定します。省略時は、"p" が設定され
ます。

a :

ログ取得モードで実行します。
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し
ます。インデックスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデックス拡張処
理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデックス拡張処理前およびインデク
ス拡張処理後にバックアップを取得する必要があります。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリ
アを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリア
が仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定でき
ません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を
参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) によ
る排他制御をするので、ほかのトランザクションと同時に実行できません。

3.3.8 phxstartidx (インデクス作成開始ユティリティ)

(1) 機能

インデクス作成方法に初期分割登録, または分割遅延登録を指定して作成したインデクスに対して, インデクス未登録状態の文書をインデクスに登録します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxstartidx -d インデクス識別子 [-t インデクス作成実行時間] [-l ログ取得方式] [-c 登録件数] [-w 同時実行指定] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インクリメンタル対象となるインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は, 英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。なお, sh, csh および ksh の場合は, さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は, 英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また, インデクス識別子にスペースが含まれる場合は, " で囲む必要があります。

-t インデクス作成実行時間

インデクス作成開始ユティリティの実行時間を分単位で指定します。

指定範囲は, 0 ~ 6,000 の間で指定してください。

0 を指定した場合, またはこのオペランドの指定を省略した場合は, 時間では制限されません。

また, 指定を省略した場合は, 実行の可否を問い合わせるメッセージを表示します。なお, 分割遅延登録・更新系 SQL 同時実行機能を利用する場合, このオプションは指定できません。

-t 指定によるインデクス作成実行時間指定は UCS4 インデクス (INDEX_UCS4 指定) の場合は使用できません。

-l ログ取得方式

インデクス登録処理でのログ取得方式を指定します。省略時は, "p" が設定されま

3. Text Search Plug-in のユティリティ

す。
分割遅延登録・更新系 SQL 同時実行機能を利用する場合は、"a" または "p" を指定してください。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデクス登録処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデクス登録処理前およびインデクス登録処理後にバックアップを取得する必要があります。

-c 登録件数

インデクス作成をする件数を、1 ~ 2,147,483,646 の範囲で指定します。
指定を省略した場合は、登録件数では制限されません。

-w 同時実行指定

このユティリティと検索処理を同時に実行するかどうか、またはこのユティリティと遅延登録を同時に実行するかどうかを指定します。省略時は、"wait" が設定されます。

- wait : 同時実行しません。
- nowait : 登録・検索を同時実行します。
- nowait_with_update : 登録・検索同時実行に加えて、分割遅延登録・更新系 SQL を同時実行します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

インデクス未登録状態の文書が存在しないインデクスに対して、このユティリティを実行しても、無効となり正常終了します。

インデクス作成実行時間を指定してインデクス作成開始ユティリティを実行した場合、インデクス未登録状態の文書数が 0 になるまでこのユティリティを実行してください。

インデクス未登録文書数は、「phxidxls (インデクス情報の取得ユティリティ)」で確認してください。

-t オプション、-c オプションの指定を省略した場合は、実行の可否を確認するメッセージを表示します。

インデクス作成開始ユティリティは、データの整合性を確保するため、インデクス作成実行時間に指定した時間が経過しても、終了しない場合があります。この場合は、データの整合性を確保できた時点で終了します。インデクス作成実行時間が経過してからユティリティが終了するまでの時間は、IXFREEWORD プラグインの環境定義ファイルに指定した一括登録時最大メモリ所要量とディスク性能に依存します。

インデクス作成実行時間を指定した場合、指定時間を過ぎてインデクス作成処理を中断して終了した場合は、メッセージログに「02001-I phxstartidx is Time Out.」のメッセージを出力します。インデクス作成処理を完了して終了した場合は、「02003-I phxstartidx command ended.」のメッセージを出力します。

-t オプションのインデクス実行時間と -c オプションの登録件数を指定した場合、どちらかの条件が成立した時点で登録処理を終了します。

同時実行指定で "wait" を指定した場合、対象処理となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。同時実行指定で "nowait" を指定した場合は、サーバ側定義が同時実行である場合にだけ有効です。

分割した表のインデクスに対し、-c オプションを指定してインデクスを作成する場合、指定した文書数のインデクスを分割した表ごとに作成します。

インデクス作成開始ユティリティは、登録要求件数分の行に対して参照できる共用排他を実行します。このため、分割遅延登録・更新系 SQL 同時実行を指定してこのユティリティを実行する場合は、排他ができるように排他資源を確保してください。

-w オプションに nowait_with_update を指定し、さらに -t オプションを指定した場合、メッセージログに「62014-E Option argument invalid. option=xx」のメッセージを出力します。

使用している HiRDB のバージョンが、この製品が前提としている HiRDB のバー

3. Text Search Plug-in のユティリティ

ジョンと不一致の場合は、エラーメッセージ「KFPL28000-E Internal error occurred」が出力され、異常終了します。

3.3.9 phxdelay (インデクス分割遅延登録運用変更ユーティリティ)

(1) 機能

インデクス作成方法に分割遅延登録を指定して作成したインデクスに対して、分割遅延登録運用開始および終了を指定します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxdelay -d インデクス識別子 -m { on|off} [-l ログ取得方式] [-f] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

分割遅延登録運用方法を変更するインデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-m

- on : インデクス分割遅延登録運用を開始します。
- off : インデクス分割遅延登録運用を終了します。

-l ログ取得方式

分割遅延登録運用方法変更インデクス拡張処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

3. Text Search Plug-in のユティリティ

更新前ログ取得モードで実行します。

ロールバックに必要なデータベース更新ログを取得します。分割遅延登録運用方法変更処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。

データベース更新ログを取得しません。分割遅延登録運用方法変更処理前および分割遅延登録運用方法変更処理後にバックアップを取得する必要があります。

-f

応答メッセージを出力しないで実行する場合に指定します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

分割遅延登録運用の終了が指定された場合、すべての未反映文書を登録します。その際、時間 / 件数による中断はできないので、分割遅延登録運用の終了以前に phxstartidx を使用して登録処理を完了させておいてください。また、分割遅延登録運用の終了実行時には、応答のメッセージを出力します。-f が指定された場合は、応答のメッセージを出力しません。

分割遅延登録用でないインデクスに対して、このユティリティを実行しても無効となります (エラーになりません。正常終了します)。

分割遅延登録・更新系 SQL 同時実行を使用している場合は、変更できません。

3.3.10 phxmodidx (インデクス情報変更ユティリティ)

(1) 機能

作成済みのインデクスに対して定義を変更します。インデクス分割遅延作成ができるインデクスへ移行できます。また、インデクス分割遅延作成用ファイルサイズを拡張できます。UCS4 インデクスへの移行を行いません。

また、UCS2 用インデクスから UCS4 用インデクスへの移行を行うこともできます。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxmodidx -d インデクス識別子 -v 変更項目 [-l ログ取得方式] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

インデクス情報変更の対象となる、インデクス識別子を指定します。必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh、csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-v 変更項目

変更するインデクス定義項目を、次の形式で指定します。

- "DELAY_SIZE= インデクス分割遅延登録用領域サイズ"

サイズは、2,608 ~ 129,584 キロバイトの範囲で指定してください。

新規に移行する際には、指定されたサイズでファイルを作成します。すでにファイルがある場合は、指定されたサイズとなるようにファイルを拡張します。最小拡張単位は 2,048 キロバイトです。端数は切り捨てられます。また、拡張回数の上限は 62 回です。

- "INDEX_TYPE = UCS4"

UCS2 用インデクスから、UCS4 用インデクスへの移行を行いません。

3. Text Search Plug-in のユティリティ

-l ログ取得方式

インデクス情報変更処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。
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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。
ロールバックに必要なデータベース更新ログを取得します。インデクス情報変更処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。
データベース更新ログを取得しません。インデクス情報変更処理前およびインデクス情報変更処理後にバックアップを取得する必要があります。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

このユティリティは、インデクス分割遅延作成用ファイルを作成または拡張します。そのまま実行すると、RD エリア不足が発生することがあります。その場合、作成または拡張するサイズ分の RD エリアを、あらかじめ追加することをお勧めします。

分割遅延登録・更新系 SQL 同時実行を使用している場合は変更できません。

このユティリティで UCS4 用インデクスへの移行を行なう場合、RD エリアの空き容量 (phnidxls コマンドの Available Work Area Size) が十分に存在していることを確

認してください。そのまま実行すると、RD エリア不足が発生することがあります。

次のデータベースに対して UCS4 用インデクスへの移行は実行できません。

- ENGLISH, ENGLISH_STANDARD (英文インデクス)
- 文字コード種別が utf-8 以外のデータベース

UCS4 インデクスへの移行時に、既にインデクス作成済み文書の UCS2 の範囲を超える UCS4 の範囲の文字は検索対象となりません。検索対象とするためには、プラグインオプションに INDEX_UCS4 を指定して、インデクスを再作成してください。

このユティリティで分割遅延作成できるインデクスへ移行しても、分割遅延登録運用の開始 / 終了状態は変更されません。運用方法を変更する場合は別途 phxdelay コマンドを実行する必要があります。また、現在の運用方法は phxidxls コマンド実行結果の「Delay Status」で確認できます

分割遅延登録・更新系 SQL 同時実行を使用している場合は、変更できません。

インデクスの再初期化を行った場合、このユティリティでの移行情報は無効になり、初めにインデクスを定義した状態に戻ります。

3.3.11 phxcond (インデクス詰め替えユティリティ)

(1) 機能

不要なインデクスを削除します。

(2) 実行者

HiRDB 管理者が実行できます。

(3) 形式

```
phxcond -d インデクス識別子 -r RDエリア名 [ -l ログ取得方式] [ -t 詰め替え時間] [-q 世代番号]
```

(4) コマンド引数

-d インデクス識別子

詰め替え対象となるインデクス識別子を指定します。

必ずインデクス作成時に指定したインデクス識別子を指定してください。

UNIX の場合

インデクス識別子を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。

また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

インデクス識別子を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。

¥" で囲まない場合はすべて大文字として扱われます。また、インデクス識別子にスペースが含まれる場合は、" で囲む必要があります。

-r RD エリア名

指定されたインデクスがある RD エリア名を指定します。

UNIX の場合

RD エリア名を " (ダブルクォーテーション) で囲んだ場合は、英大文字と英小文字が区別されます。" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。なお、sh, csh および ksh の場合は、さらに ' (アポストロフィ) で囲む必要があります。

Windows の場合

RD エリア名を ¥" で囲んだ場合は、英大文字と英小文字が区別されます。¥" で囲まない場合はすべて大文字として扱われます。また、RD エリア名にスペースが含まれる場合は、" で囲む必要があります。

-l ログ取得方式

インデクス詰め替え処理でのログ取得方式を指定します。省略時は、"p" が設定されます。

a :

ログ取得モードで実行します。

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します。インデクスのバックアップを取得する必要はありません。

p :

更新前ログ取得モードで実行します。

ロールバックに必要なデータベース更新ログを取得します。インデクス詰め替え処理後、バックアップを取得する必要があります。

n :

ログレスモードで実行します。

データベース更新ログを取得しません。インデクス詰め替え処理前およびインデクス詰め替え処理後にバックアップを取得する必要があります。

-t 詰め替え時間

詰め替えユティリティの実行時間を分単位で指定します。

実行時間は半角数字で 0 ~ 6,000 の間で 15 分単位で指定してください。15 の倍数でない値が指定された場合は、15 分単位で切り上げられます。

実行時間に 0 を指定した場合、またはこのオペランドを省略した場合、インクリメンタル処理が完了するまで処理を続けます。

指定した実行時間でインクリメンタル処理が完了しなかった場合、ユティリティが指定時間を過ぎて終了した旨のメッセージを出力します。

-q 世代番号

Staticizer Option のインナレプリカ機能を使用する場合、処理対象とする RD エリアを 0 ~ 10 の世代番号で指定します。指定を省略した場合は、カレント RD エリアが仮定されます。

- 0 : オリジナル RD エリアを対象とします。
- 1 ~ 10 : 指定した世代番号の RD エリアを対象とします。

このオペランドは、Staticizer Option がインストールされていない場合は指定できません。

なお、インナレプリカ機能については、マニュアル「HiRDB Staticizer Option」を参照してください。

(5) 規則

このユティリティは、HiRDB が稼働中のときだけ実行できます。

(6) 注意事項

詰め替え処理は、1 回の処理ですべてのインデクスを処理できないことがあります。

3. Text Search Plug-in のユティリティ

このため、詰め替え実施後、「phxidxls (インデクス情報の取得ユティリティ)」で未実施インデクス数 (Number of No Condensed Index) を確認してください。この値が大ききときは、インデクス空き容量に注意して、再度詰め替えを実行してください。

このユティリティは、インデクス空き容量がある状態 (推奨 10% 以上) で実行する必要があります。空きのない状態でユティリティを実行したとき、または詰め替え未実施インデクス数 (Number of No Condensed Index) が減らない場合は、新たに RD エリアを追加して「phxexpnd (インデクスの拡張ユティリティ)」を実行し、空き容量がある状態で実行してください。空き容量は、「phxidxls (インデクス情報の取得ユティリティ)」の Free Size of Index File で確認してください。

詰め替え実行時間を指定した場合、指定時間と実際の処理時間との間に多少のずれが生じることがあります。

文書を削除すると、詰め替え可能文書率は増加します。このため、詰め替え中断時に文書を削除すると、詰め替え処理完了後も詰め替え可能文書率は 0 になりません。

このユティリティ実行中は、処理対象となる RD エリアに対して排他モード (EX) による排他制御をするので、ほかのトランザクションと同時に実行できません。

詰め替えを実行する契機の指標として、「phxidxls (インデクス情報の取得ユティリティ)」の「Ratio of Condensable Documents (詰め替え可能文書率)」が、50% 以上を目安としてください。

「phxidxls (インデクス情報の取得ユティリティ)」の「Ratio of Condensed Index (詰め替え作業率)」は、詰め替え処理が制限時間で終了したときの詰め替え作業進捗率を表します。

このユティリティを実行する際に、グローバルバッファを指定すると、実行性能が向上します。

詰め替え時間を指定した場合、指定時間を過ぎて詰め替え処理を中断したときは、メッセージログに「02001-I phxcond is Time Out.」のメッセージが出力されます。詰め替え処理を完了して終了した場合は、「02003-I phxcond command ended」のメッセージが出力されます。

4

抽象データ型関数を使用した文書または文字列データの操作

この章では、文書または文字列データを操作するために Text Search Plug-in が提供している、抽象データ型関数の使用方法および検索条件の文法について説明します。

4.1 Text Search Plug-in が提供する抽象データ型関数

4.2 抽象データ型関数の文法

4.3 全文検索条件式の文法（文書検索プラグインの場合）

4.4 全文検索条件式の文法（文字列検索プラグインの場合）

4.5 概念検索条件式の文法

4.1 Text Search Plug-in が提供する抽象データ型関数

この節では、Text Search Plug-in が提供する抽象データ型関数の使用方法について説明します。

4.1.1 抽象データ型関数の使用方法

Text Search Plug-in の抽象データ型関数とは、Text Search Plug-in が提供する全文検索機能および概念検索機能を利用するためのインタフェース関数です。文書検索プラグインは、SGMLTEXT 型のデータ（文書）に対して全文検索および概念検索を利用するために、HiRDB に SGMLTEXT 型を追加します。また、文字列検索プラグインは、FREEWORD 型のデータ（文字列データ）に対して全文検索を利用するために、HiRDB に FREEWORD 型を追加します。

SGMLTEXT 型で全文検索および概念検索をする場合は、抽象データ型関数 SGMLTEXT を使用して、表に SGMLTEXT 型のデータとして文書を登録します。

FREEWORD 型で全文検索をする場合は、抽象データ型関数 FREEWORD を使用して、表に FREEWORD 型のデータとして文字列データを登録します。

なお、FREEWORD 型では、概念検索機能は使用できません。

(1) 抽象データ型関数の種類

Text Search Plug-in の提供する抽象データ型関数の種類を次に示します。

文書検索プラグインの場合

SGMLTEXT 型のデータ（文書）に対するインタフェースとして、次の 3 種類の抽象データ型関数を提供しています。

- データ出力用の抽象データ型関数
- 全文検索用の抽象データ型関数
- 概念検索用の抽象データ型関数

文字列検索プラグインの場合

FREEWORD 型のデータ（文字列データ）に対するインタフェースとして、次の 2 種類の抽象データ型関数を提供しています。

- データ出力の抽象データ型関数
- 全文検索用の抽象データ型関数

上記のうち、SGMLTEXT 型の全文検索用の抽象データ型関数は、SGMLTEXT 型の全文検索用インデクスおよび概念検索用インデクスに対して使用できます。SGMLTEXT 型の概念検索用の抽象データ型関数は、SGMLTEXT 型の概念検索用インデクスに対してだけ使用できます。また、FREEWORD 型の全文検索用の抽象データ型関数は、

FREEWORD 型の全文検索用インデクスに対して使用できます。これらの抽象データ型関数を SQL 文中に記述することで、SGMLTEXT 型のデータ（文書）または FREEWORD 型のデータ（文字列データ）にアクセスできます。SQL の記述形式については、マニュアル「HiRDB SQL リファレンス」を参照してください。

(2) Text Search Plug-in が提供する抽象データ型関数の一覧

Text Search Plug-in が提供する抽象データ型関数の一覧を表 4-1 に示します。

表 4-1 Text Search Plug-in の抽象データ型関数一覧

| 用途 | 関数名 | 機能 | 型別サポート有無 | |
|--------|---------------------|---|-------------|-------------|
| | | | SGML TEXT 型 | FREE WORD 型 |
| データ登録用 | SGMLTEXT | SGMLTEXT 型のインスタンスを生成するコンストラクタ関数です。SGMLTEXT 型の列にデータ（文書）を登録する場合は、この関数を使用してください。 | | × |
| | FREEWORD | FREEWORD 型のインスタンスを生成するコンストラクタ関数です。FREEWORD 型の列にデータ（文字列データ）を登録する場合は、この関数を使用してください。 | × | |
| データ出力用 | extracts | 登録した文書または文字列データを出力します。出力する形式には、次の 4 種類あります。 | | |
| | | • 原文書出力 | | |
| | | • SGML 出力（表示用 DTD 付加） | | × |
| | | • プレーン出力 | | × |
| | | • XML 出力 | | × |
| 全文検索用 | contains | 全文検索条件式に合致する文書または文字列データを検索します。全文検索条件式には、次の 6 種類あります。 | | |
| | | • 単純文字列指定 | | |
| | | • 異表記展開指定 | | |
| | | • 同義語展開指定 | | |
| | | • 近傍条件指定 | | |
| | | • AND/OR 論理条件 | | |
| | • 構造名指定による検索範囲の指定 | | × | |
| | contains_with_score | 全文検索条件式に合致する文書を検索します。 | | |

4. 抽象データ型関数を使用した文書または文字列データの操作

| 用途 | 関数名 | 機能 | 型別サポート有無 | |
|-------|--------------------|--|-------------------|------------------------|
| | | | SGML TEXT 型 | FREE WORD D 型 |
| | score | contains_with_score 関数で検索した文書のスコア値を取得します。検索条件に合致した文書ほどスコア値が高くなります。 | | |
| 概念検索用 | concept_with_score | 概念検索条件式に適合する文書を検索します。また、スコアの計算をします。 | | × |
| | score_concept | concept_with_score 関数で検索した文書のスコア値を取得します。検索条件に対する適合度が高い文書ほどスコア値が高くなります。 | | × |
| | concept_with_terms | 概念検索条件式として指定した種文章から検索用特徴タームと重みを抽出します。 | | × |
| | terms_concept | concept_with_terms 関数で抽出した検索用特徴タームと重みを取得します。 | | × |

(凡例)

○ : サポートあり

× : サポートなし

4.1.2 抽象データ型の選択

ここでは、Text Search Plug-in が提供する抽象データ型の選択方法について説明します。

(1) 提供する抽象データ型の相違点

Text Search Plug-in が提供する抽象データ型として SGMLTEXT 型と FREEWORD 型があります。

これらの抽象データ型の相違点を表 4-2 に示します。表 4-2 を参照して使用する抽象データ型を選択してください。

表 4-2 抽象データ型の相違点

| 項目 | SGMLTEXT 型 | FREEWORD 型 |
|-------------|--|--------------------------|
| 登録できるデータの種別 | 構造化文書 <ul style="list-style-type: none"> • SGML • XML (Preprocessing Library for Text Search で作成したテキストデータ) | プレーン文書 (文字列データ) |
| 登録データ長の上限值 | 5 メガバイト | 32,000 バイト (推奨 255 バイト) |

| 項目 | SGMLTEXT 型 | FREWORD 型 |
|-----------------|--|-----------------------|
| 登録データ格納用 RD エリア | ユーザ用 RD エリア (基表) ユーザ LOB 用 RD エリア (データ) | ユーザ用 RD エリア (基表, データ) |
| 全文検索 | | |
| 概念検索 | | × |
| 登録データの出力 | 原文書出力 SGML 出力 プレーン出力 XML 出力 | 原文書出力 |
| 対応するインデクス型 | NGRAM 型 | IXFREEWORD 型 |
| インデクス格納用 RD エリア | ユーザ LOB 用 RD エリア | ユーザ LOB 用 RD エリア |
| 差分インデクス定義 | | × |
| インデクス作成ルール追加機能 | | × |
| インデクスファイル自動拡張機能 | | (設定を推奨します) |

(凡例)

- : 使用できます
- × : 使用できません

注

構造名を指定した検索はできません。

(2) 非構造文書を扱う場合に推奨する抽象データ型

構造を持たない文書 (または文字列データ) は, SGMLTEXT 型 (表の作成時に指定する文書登録種別にプレーン文書を指定した場合) と FREEWORD 型の両者に登録できます。ただし, 次の条件を満たす場合, FREEWORD 型にデータを登録することを推奨します。

登録データ長が 3 キロバイト以下の文書または文字列データを扱う場合

全文検索のスコア値の取得や概念検索を使用しない場合

4.2 抽象データ型関数の文法

この節では、Text Search Plug-in の抽象データ型関数の詳細について説明します。なお、各関数で指定する全文検索条件式の文法の詳細については、文書検索プラグインの場合は、「4.3 全文検索条件式の文法（文書検索プラグインの場合）」、文字列検索プラグインの場合は、「4.4 全文検索条件式の文法（文字列検索プラグインの場合）」を、概念検索条件式の文法の詳細については「4.5 概念検索条件式の文法」をそれぞれ参照してください。

4.2.1 SGMLTEXT

(1) 機能

SGMLTEXT 型のインスタンスを生成します。

(2) 書式

SGMLTEXT (文書データ)

(3) 入力

文書データ (BLOB)

(4) 戻り値

SGMLTEXT 型インスタンス

(5) 詳細

文書データには、テキスト形式の SGML 文書またはプレーン文書を BLOB 型変数に入れて指定します。

扱える文書サイズは最大 5 メガバイトです。

0 バイトの文書は文書検索プラグインに登録できません。文書検索プラグインに 0 バイトの文書に登録するときは、ナル値で登録してください。

(6) SQL 文の例

```
EXEC SQL BEGIN DECLARE SECTION;
      SQL TYPE IS BLOB(1M) sgml;          BLOB型変数の定義
EXEC SQL END DECLARE SECTION;

strcpy(sgml.sgml_data,                   BLOB型変数への代入
      char_ptr_pointing_to_a_sgml_text);

sgml.sgml_length =                        BLOB型変数データ長の代入
      strlen(char_ptr_pointing_to_a_sgml_text);

EXEC SQL                                INSERT文の実行
```


4. 抽象データ型関数を使用した文書または文字列データの操作

```
INSERT INTO reports(doc) VALUES(SGMLTEXT(:sgml AS BLOB(1M)));
```

4.2.2 FREEWORD

(1) 機能

FREEWORD 型のインスタンスを生成します。

(2) 書式

FREEWORD (文字列データ)

(3) 入力

文字列データ (VARCHAR)

(4) 戻り値

FREEWORD 型インスタンス

(5) 詳細

文字列データには、テキスト形式の文字列を VARCHAR 型変数に入れて指定します。

扱える文字列サイズは最大 32,000 バイトです。

(6) SQL 文の例

```
EXEC SQL BEGIN DECLARE SECTION;
    struct {
        short length;
        char data[32000];
    }string;
EXEC SQL END DECLARE SECTION;

strcpy(string.data,
    char_ptr_pointing_to_a_fword_string);

string.length =
    strlen(char_ptr_pointing_to_a_fword_string);

EXEC SQL
    INSERT INTO reports(fword) VALUES(FREEWORD(:string AS
    VARCHAR(32000)));
```

4.2.3 contains

(1) 機能

全文検索条件式に基づき検索します。

(2) 書式

contains (列指定, 全文検索条件式)

(3) 入力

列指定

全文検索条件式 (varchar) (最大 32,000 バイト)

(4) 戻り値

常に TRUE

(5) 詳細

列指定には、全文検索をする SGMLTEXT 型または FREEWORD 型の列名を指定します。

全文検索条件式には、検索する条件として検索対象の文書構造名、検索タームなどを指定した条件式 (varchar) を指定します。

contains 関数はインデクスと協調して動作する検索述語であり、contains 関数使用時にインデクス型プラグイン (SGML プラグインに対して n-gram インデクスプラグイン、FREEWORD プラグインに対して IXFREEWORD プラグイン) が使用できない状態の場合には、エラーとなります。インデクス型プラグイン専用関数の実行方法に関する規則については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書の件数を調べる (文書の最上位構造は「文章」)。

```
SELECT COUNT(*) FROM reports
WHERE contains(doc, '文章[概要{"COMPUTER"}]') IS TRUE
```

表「reports」の FREEWORD 型の列「fword」に文字列「COMPUTER」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{"COMPUTER"}') IS TRUE
```

(7) 注意事項

全文検索条件式には、複数の条件を "AND", "OR" で指定できます。

4.2.4 contains_with_score

(1) 機能

全文検索条件式に基づき検索します。また、同時にスコア値の計算をします。

(2) 書式

contains_with_score (列指定, 全文検索条件式)

(3) 入力

列指定

全文検索条件式 (varchar) (最大 32,000 バイト)

(4) 戻り値

常に TRUE

(5) 詳細

列指定には全文検索をする SGMLTEXT 型, または FREEWORD 型の列名を指定します。

全文検索条件式には, 検索する条件として検索対象の文書構造名, 検索タームなどを指定した条件式 (varchar) を指定します。

contains_with_score 関数は, インデクスと協調して動作する検索述語です。contains_with_score 関数使用時にインデクス型プラグイン (SGML プラグインに対して n-gram インデクスプラグイン, FREEWORD プラグインに対して IXFREEWORD インデクスプラグイン) が使用できない状態の場合には, エラーとなります。

contains_with_score 関数はスコア値を算出し, score 関数に値を受け渡す関数です (受け渡し値送信関数)。スコア値は, score 関数で受け取ります (受け渡し値受信関数)。インデクス型プラグイン専用関数および受け渡し値送受信関数の実行方法に関する規則については, マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書をスコア値順にソートし, スコア値, タイトルを格納する一覧表を作成する (文書の最上位構造は「文章」)。

```
SELECT score(doc), title FROM reports
WHERE contains_with_score(doc, '文章[概要{"COMPUTER"}]' ) IS
TRUE
ORDER BY 1 DESC
```

(7) 注意事項

全文検索条件式には、複数の条件を "AND", "OR" で指定できます。

4.2.5 score

(1) 機能

スコア値 (contains_with_score 関数の結果として得られるスコア値) を返却します。

(2) 書式

score (列指定)

(3) 入力

列指定

(4) 戻り値

スコア値 (INTEGER)

(5) 詳細

contains_with_score 関数と一緒に使用します。

列指定は contains_with_score 関数で検索する対象と同じ列名を指定します。score 関数を使用する SELECT では、必ず WHERE 探索条件に score 関数と同じ列名を指定した contains_with_score 関数を使用します。検索結果をスコア順にソートして、スコアの得点の高いものから取り出す場合には、ORDER BY を使用してください。

score 関数は受け渡し値受信関数であり、contains_with_score 関数から値を受け取ります。受け渡し値受信関数の実行方法に関する規則については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) スコア値

1 文書のスコア値は、「検索タームの文書内での出現数 × 検索タームの重み × 構造の重み」で算出されます。ただし、検索条件式に論理検索や近傍検索を使用している場合は、次のようになります。

論理検索の場合

- 論理積 (AND) のときは、各スコアの最小値を取ります。
- 論理和 (OR) のときは、各スコアの最大値を取ります。
- 論理否定 (NOT) のときは、各スコア値はナル値となります。

近傍検索の場合

- 近傍検索のスコアは、ヒットした検索ターム間の距離で算出します。
- 検索ターム間の距離が近いほどスコアが高くなります。

表 4-3 スコアの算出方法

| 検索条件 | スコアの算出方法 | 算出例 |
|------------------------------|----------------------------|--|
| 論理積 (AND 条件) ¹ | 論理演算の対象を比較して、値の少ない方をスコアとする | 「A および B が含まれる文書」を検索した結果、文書内での出現数が A>B だった場合、B の出現数を基にスコアが付けられる ² |
| 論理和 (OR 条件) ¹ | 論理演算の対象を比較して、値の多い方をスコアとする | 「A または B が含まれる文書」を検索した結果、文書内での出現数が A>B だった場合、A の出現数を基にスコアが付けられる ² |
| 近傍条件 | = | 指定した距離とヒットした距離が等しい場合は 100 が与えられる |
| | <= | ヒットした距離が 0 に近いほど 100 に近づく。指定した距離と同じ場合は 1 が与えられる |
| | >= | 指定した距離とヒットした距離が等しい場合は 100 が与えられる。遠ざかるほど低くなる |
| | | 「AB 間の距離が 10 文字ちょうどの文書」を検索した結果、文書内での A と B の距離が 10 であれば 100 が与えられる |
| | | 「AB 間の距離が 10 文字以下の文書」を検索した結果、文書内での A と B の距離が 0 ならば 100 が与えられる。10 以下の値は、10 に近いほど低い値が与えられる |
| | | 「AB 間の距離が 10 文字以上の文書」を検索した結果、文書内での A と B の距離が 10 ならば 100 が与えられる。10 を超える値は、10 から遠くなるほど低いスコアが与えられる |

注 1

検索ターム同士および検索条件式同士の論理演算が含まれます。

注 2

ここでの算出方法は、検索条件式の構造に重みが付けられていないことが前提です。構造に重みが付けられている場合は、「距離から算出したスコア × 構造の重み」で算出されます。なお、近傍検索では検索タームに付けられた重みは無視されます。

4.2.6 extracts

登録した文書または文字列データを出力します。出力形式には次に示す種類があります。

原文書出力 (SGML 型と FREEWORD 型で形式が異なります)

SGML 出力

プレーン出力

XML 出力

どの出力形式で出力するかは、extracts 関数の引数で区別されます。

なお、FREEWORD 型では、SGML 出力、プレーン出力、および XML 出力は使用できません。

特に、文字コード utf-8 でこの関数を使用する場合、次の制限があります。

SGML 出力は使用できません。

引数は必ず MS-Unicode で記述してください。

また、extracts 関数を使用した SQL の前処理後 (PREPARE 文の発行後)、DESCRIBE [OUTPUT] 文の実行時に SQL 記述領域に設定されるデータの長さは次のとおりです。

SGMLTEXT 型の場合：2 ギガバイト (2,147,483,647 バイト)

FREEWORD 型の場合：32,000 バイト

次に、各出力形式について説明します。

(1) 原文書出力 (SGMLTEXT 型の場合)

(a) 機能

登録した文書全体 (原文書) を出力します。

(b) 書式

extracts (列指定)

(c) 入力

列指定

(d) 戻り値

テキストデータ (BLOB)

(e) 詳細

列指定には、検索をする SGMLTEXT 型の列名を指定します。

(f) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書全体（原文書）を出力する（文書の最上位構造は「文章」）。

```
SELECT extracts(doc) FROM reports
WHERE contains(doc, '文章[概要{"COMPUTER"}]' ) IS TRUE
```

(2) 原文書出力（FREEWORD 型の場合）

(a) 機能

登録した文字列データ全体（原文書）を出力します。

(b) 書式

extracts (列指定)

(c) 入力

列指定

(d) 戻り値

文字列データ (VARCHAR)

(e) 詳細

列指定には、FREEWORD 型の列名称を指定します。

(f) SQL 文の例

表「reports」の FREEWORD 型の列「fword」に文字列「COMPUTER」を含む文字列全体（原文書）を出力する。

```
SELECT extracts(fword) FROM reports
WHERE contains(fword, '{"COMPUTER"}' ) IS TRUE
```

(3) SGML 出力

(a) 機能

SGML 文書の特定構造の抽出およびハイライトタグの埋め込みをして、SGML 形式で出力します。出力文書には、DTD を付けます。

(b) 書式

extracts (列指定, 抽出構造名, ハイライト位置条件式, ハイライトタグ名)

(c) 入力

列指定

抽出構造名 (varchar) (最大 1,024 バイト)

ハイライト位置条件式 (varchar) (最大 32,000 バイト)

4. 抽象データ型関数を使用した文書または文字列データの操作

ハイライトタグ名 (varchar)(最大 255 バイト)

(d) 戻り値

テキストデータ (BLOB)

(e) 詳細

列指定には、検索する SGMLTEXT 型の列名を指定します。なお、FREEWORD 型の列は指定できません。

抽出構造名には、抽出する文書要素のその構造名を指定します。何も指定しない場合は、文書全体を出力します。下位の構造を指定する場合、最上位の構造から順に記述し、"." (ピリオド) でつないでください。

複数個の文書要素を抽出する場合は、抽出構造名を ";" (セミコロン) で区切って記述してください。なお、抽出構造名はワイルドカードが指定できます。ワイルドカード指定の方法は、全文検索条件式の構造名と同形式です。

ハイライト位置条件式には、ハイライトタグを埋め込む位置を指定する条件式を指定します。ハイライト位置条件式は、全文検索条件式と同形式です。ただし、重みおよび NOT 条件は記述しても無視されます。除外文字の検索は指定できません。何も指定しない場合はハイライトタグは埋め込まれません。

ハイライトタグ名には、ハイライト位置条件式を満たす位置に埋め込むタグの名称を指定します。何も指定しない場合にはハイライトタグは埋め込まれません。

(f) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書の「概要」を SGML 形式で出力する。その際、文字列「COMPUTER」を強調表示するハイライトタグ (STRONG) を埋め込む (文書の最上位構造は「文章」)。

```
SELECT extracts(doc, '文章.概要', '文章[概要{"COMPUTER"}]',  
               'STRONG')  
FROM reports  
WHERE contains(doc, '文章[概要{"COMPUTER"}]') IS TRUE
```

(4) プレーン出力

(a) 機能

構造化文書の特定構造の抽出をして、プレーン形式で出力します。

(b) 書式

extracts (列指定, 抽出構造名)

(c) 入力

列指定

抽出構造名 (varchar)(最大 1,024 バイト)

(d) 戻り値

テキストデータ (BLOB)

(e) 詳細

列指定には、検索する SGMLTEXT 型の列名を指定します。なお、FREEWORD 型の列は指定できません。

抽出構造名には、抽出する文書要素のその構造名で指定します。何も指定しない場合は、文書全体を出力します。下位の構造を指定する場合、最上位の構造から順に記述し、"." (ピリオド) でつないでください。

複数個の文書要素を抽出する場合は、抽出構造名を ";" (セミコロン) で区切って記述してください。なお、抽出構造名はワイルドカードが指定できます。ワイルドカード指定の方法は、全文検索条件の構造名と同形式です。

(f) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書の「概要」をプレーン形式で出力する (文書の最上位構造は「文章」)。

```
SELECT extracts(doc, '文章.概要')
FROM reports
WHERE contains(doc, '文章[概要{"COMPUTER"}]') IS TRUE
```

(5) XML 出力

(a) 機能

構造化文書の特定制造の抽出、およびハイライトタグの埋め込みをして、ウェルフォームド XML 文書で出力します。出力文書には、XML 宣言を付けます。

(b) 書式

extracts (列指定, 抽出構造名, ハイライト位置条件式, ハイライトタグ名, ドキュメントタイプ)

(c) 入力

列指定

抽出構造名 (varchar) (最大 1,024 バイト)

ハイライト位置条件式 (varchar) (最大 32,000 バイト)

ハイライトタグ名 (varchar) (最大 255 バイト)

ドキュメントタイプ 'XML'

(d) 戻り値

テキストデータ (BLOB)

4. 抽象データ型関数を使用した文書または文字列データの操作

(e) 詳細

列指定には、検索する SGMLTEXT 型の列名を指定します。なお、FREEWORD 型の列は指定できません。

抽出構造名には、抽出する文書要素のその構造名を指定します。何も指定しない場合は、文書全体を出力します。下位の構造を指定する場合、最上位の構造から順に記述し、"." (ピリオド) でつないでください。

複数個の文書要素を抽出する場合は、抽出構造名を ";" (セミコロン) で区切って記述してください。なお、抽出構造名はワイルドカードが指定できます。ワイルドカード指定の方法は、全文検索条件式の構造名と同形式です。

ハイライト位置条件式には、ハイライトタグを埋め込む位置を指定する条件式を指定します。ハイライト位置条件式は、全文検索条件式と同形式です。ただし、重みおよび NOT 条件は記述しても無視されます。除外文字検索は指定できません。何も指定しない場合はハイライトタグは埋め込まれません。

ハイライトタグ名には、ハイライト位置条件式を満たす位置に埋め込むタグの名称を指定します。何も指定しない場合にはハイライトタグは埋め込まれません。

ドキュメントタイプには 'XML' を指定してください。

出力される XML 文書は、XML 規格のバージョン 1.0 のウェルフォームド XML 文書です。

(f) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分に文字列「COMPUTER」を含む文書の「概要」を XML 形式で出力する。その際、文字列「COMPUTER」を強調表示するハイライトタグ (STRONG) を埋め込む (文書の最上位構造は「文章」)。

```
SELECT extracts(doc, '文章.概要', '文章[概要{"COMPUTER"}]',
  'STRONG', 'XML')
  FROM reports
 WHERE contains(doc, '文章[概要{"COMPUTER"}]') IS TRUE
```

4.2.7 concept_with_score

(1) 機能

概念検索条件式に基づき検索します。また、同時にスコア値の計算をします。

(2) 書式

concept_with_score (列指定, 概念検索条件式)

(3) 入力

列指定

概念検索条件式 (BLOB) (最大 5 メガバイト)

(4) 戻り値

常に TRUE

(5) 詳細

列指定には、概念検索をする SGMLTEXT 型の列名称を指定します。なお、FREEWORD 型の指定はできません。

概念検索条件式には、検索する条件としての文章 (種文章) および検索対象の文書構造名を指定した条件式 (BLOB) を指定します。

concept_with_score 関数は、n-gram インデクスと協調して動作する検索述語です。concept_with_score 関数使用時にインデクス型プラグイン (n-gram インデクスプラグイン) が使用できない状態の場合には、エラーとなります。

concept_with_score 関数はスコア値を算出し、score_concept 関数に値を受け渡す関数です (受け渡し値送信関数)。スコア値は、score_concept 関数で受け取ります (受け渡し値受信関数)。インデクス型プラグイン専用関数および受け渡し値送受信関数の実行方法に関する規則については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」の「概要」部分について、種文章と似た概念を持つ文書を検索する。その後、スコア順にソートし、スコア値、タイトルを格納する一覧表を作成する (文書の最上位構造は「文章」)。

```
EXEC SQL BEGIN DECLARE SECTION;
      struct {long len; char data[4096];} sql;
      SQL TYPE IS BLOB(1M) excond;
      char title[8];
      int score;
EXEC SQL END DECLARE SECTION;
```

4. 抽象データ型関数を使用した文書または文字列データの操作

```
strcpy(sql.data, "SELECT score_concept(doc), title FROM reports
WHERE concept_with_score(doc, ? AS BLOB(1M)) IS TRUE ORDER
BY 1 DESC");
sql.len = strlen(sql.data);

strcpy(excond.excond_data,
"文章 [概要{¥}概念検索とは, 検索条件として入力された文章と
内容(概念)¥が似通った文書を検索する機能である。¥"}]");
excond.excond_length = strlen(excond.excond_data);
EXEC SQL PREPARE sqlxt FROM :sql;
EXEC SQL DECLARE excrs CURSOR FOR sqlxt;
EXEC SQL OPEN excrs USING :excond;
while(1) {
    EXEC SQL FETCH excrs INTO :score, :title;
}
score_concept
```

(7) 注意事項

全文検索条件式には, 複数の条件を "AND", "OR" で指定できます。

4.2.8 score_concept

(1) 機能

スコア値 (concept_with_score 関数の結果として得られるスコア値) を返却します。

(2) 書式

score_concept (列指定)

(3) 入力

列指定

(4) 戻り値

スコア値 (INTEGER)

(5) 詳細

concept_with_score 関数と一緒に使用します。

列指定は concept_with_score 関数で検索する対象と同じ列名を指定します。

score_concept 関数を使用する SELECT では、必ず WHERE 探索条件に score_concept 関数と同じ列名を指定した concept_with_score 関数を使用します。検索結果をスコア順にソートして、スコアの得点の高いものから取り出す場合には、ORDER BY を使用してください。

score_concept 関数は受け渡し値受信関数であり、concept_with_score 関数から値を受け取ります。受け渡し値受信関数の実行方法に関する規則については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) スコア値

1 文書のスコア値は、1 個以上の種文章から抽出された各検索用特徴タームのスコアの総和です。検索用特徴タームのスコアは、検索用特徴タームの、データベースの登録文書全体に出現する割合、種文章中に出現する割合およびヒット文書中に出現する割合から算出されます。

検索用特徴タームのスコアは、次の場合に高くなります。

登録文書全体に対する出現割合が低い場合

種文章に対する出現割合が高い場合

ヒット文書に対する出現割合が高い場合

算出方法には、次の二つの方法があります。ただし、どちらの方法でもスコア値が 2,147,483,647 を超える場合は、上限値 2,147,483,647 として扱われます。

複合論理演算での算出法

4. 抽象データ型関数を使用した文書または文字列データの操作

- スコア総和方式に基づいて算出されます。
- AND：該当文書同士のスコアの和を総スコアとします。
- OR：該当文書同士のスコアの和を総スコアとします。
- NOT：各スコア値はナル値となります。

構造指定に上位構造（下位構造を複数含む）を指定した場合の算出法

- 該当構造ごとにスコアを算出し、その総和をスコアとします。

なお、スコアは、種文章を 100 点とした相対的な値に正規化して出力することもできます。スコアの値を正規化して出力する方法については、「4.5.9 スコアオプション」を参照してください。

(7) 注意事項

全文検索条件式には、複数の条件を "AND", "OR" で指定できます。

4.2.9 concept_with_terms

(1) 機能

指定した種文章から検索用特徴タームと重みを抽出します。重みは、種文章から抽出される検索用特徴タームの中で重要度が最も高いタームを基準（100）として値が返却されます。なお、種文章から抽出する検索用特徴タームの個数の上限値は、n-gram インデックスプラグインの環境設定項目「検索用特徴ターム数」に依存します。

(2) 書式

concept_with_terms (列指定, 概念検索条件式)

(3) 入力

列指定

概念検索条件式 (BLOB) (最大 5 メガバイト)

(4) 戻り値

常に TRUE

(5) 詳細

列指定には、検索用特徴タームと重みを抽出する SGMLTEXT 型の列名を指定します。なお、FREEWORD 型の列は指定できません。

概念検索条件式には、検索する条件としての文章（種文章）を指定します。

概念検索条件式で指定できるのは単純文字列指定だけです。ほかの指定はエラーとなります。

concept_with_terms 関数は、n-gram インデックスと協調して動作する検索述語であり、concept_with_terms 関数使用時にインデックス型プラグイン（n-gram インデックスプラグイン）が使用できない状態の場合には、エラーとなります。

concept_with_terms 関数は検索用特徴タームと重みを抽出し、terms_concept 関数に値を受け渡す関数です（受け渡し値送信関数）。検索用タームと重みは、terms_concept 関数で受け取ります（受け渡し値受信関数）。インデックス型プラグイン専用関数および受け渡し値送受信関数の実行方法に関する規則については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) SQL 文の例

表「reports」の SGMLTEXT 型の列「doc」に対して、種文章と似た文書を検索するための検索用特徴タームと重みを抽出する。

```
EXEC SQL BEGIN DECLARE SECTION;
      struct {long len; char data[4096];} sql;
SQL TYPE IS BLOB(1M) excond;
```

4. 抽象データ型関数を使用した文書または文字列データの操作

```
    struct {
        short length ;
        char data[32000] ;
    }terms ;
EXEC SQL END DECLARE SECTION;

strcpy(sql.data, "SELECT terms_concept(doc) FROM reports
    WHERE concept_with_terms(doc, ? AS BLOB(1M)) IS TRUE ORDER
BY 1 DESC");
sql.len = strlen(sql.data);

strcpy(excond.excond_data,
    "{¥"概念検索とは、検索条件として入力された文章と内容(概念)¥
    が似通った文書を検索する機能である。¥"}");
excond.excond_length = strlen(excond.excond_data);
EXEC SQL PREPARE sqlxt FROM :sql;
EXEC SQL DECLARE excrs CURSOR FOR sqlxt;
EXEC SQL OPEN excrs USING :excond;
while(1) {
    EXEC SQL FETCH excrs INTO :terms;
}
```

4.2.10 terms_concept

(1) 機能

検索用特徴タームと重み (concept_with_terms 関数の結果として得られる検索用特徴タームと重み) を返却します。

(2) 書式

terms_concept (列指定)

(3) 入力

列指定

(4) 戻り値

検索用特徴タームと重み (varchar) (最大 32,000 バイト)

(5) 詳細

concept_with_terms 関数と一緒に使用します。

列指定は concept_with_terms 関数で検索する対象と同じ列名を指定します。

terms_concept 関数を使用する SELECT では、必ず WHERE 探索条件に terms_concept 関数と同じ列名を指定した concept_with_terms 関数を使用します。

terms_concept 関数は受け渡し値受信関数であり、terms_with_score 関数から値を受け取ります。受け渡し値受信関数の実行方法に関する規則は、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

(6) 検索用特徴タームと重みの取得形式

検索用特徴タームと重みの取得形式を次の図に示します。

図 4-1 検索用特徴タームと重みの取得形式

```

取得形式 ::= 検索用特徴ターム情報群 [ , 検索用特徴ターム情報群 [ , ... ] ] 終端文字
検索用特徴ターム情報群 ::=
    左波括弧 左括弧 検索用特徴ターム情報 [ , 検索用特徴ターム情報 [ , ... ] ] 右括弧 右波括弧
検索用特徴ターム情報 ::=
    {
        ダブルクォーテーション 検索用特徴ターム ダブルクォーテーション 重み
        情報なし
    }
重み ::= 左括弧 数値 右括弧
終端文字 ::= '¥0'
情報なし ::= ''

```

4. 抽象データ型関数を使用した文書または文字列データの操作

検索用特徴ターム情報群は、concept_with_terms 関数の特徴ターム取得条件式に指定した単純文字列指定に対する、検索用特徴タームおよび重みの集まりです。

concept_with_terms 関数の特徴ターム取得条件式として、二つ以上の単純文字列指定を指定した場合は、単純文字列指定ごとの検索用特徴ターム情報群が、指定した順番で、コンマで区切られて返却されます。

なお、ある単純文字列指定に対する検索用特徴タームが一つも存在しない場合は、その単純文字列に対する検索用特徴ターム情報群は空（情報なし）で返却されます。ただし、検索用特徴ターム情報群すべてが空の場合は、値は返却されません。

返却する情報が 32,000 バイトを超える場合は、32,000 バイトに格納できる検索用特徴ターム情報群までを返却します。

検索用特徴ターム情報群に含まれる検索用特徴タームは、重要度が最も高いタームを基準（100）として、重みの大きい順に返却されます。重みの値は 1 ~ 100 の整数値です。

返却する情報の終端には '¥0'（0x00）が格納されます。

（7）検索用特徴タームと重みの取得例

検索用特徴タームと重みの取得例を次に示します。

（a）一つの種文章に対して、検索用特徴タームと重みを取得した場合

概念検索条件式：

```
{(" 検索条件として入力された文章と概念が似通った文書を検索する。")}
```

出力結果：

```
{(" 検索 "(100)," 条件 "(82)," 入力 "(75)," 文章 "(55)," 概念 "(54)," 似 "(43)," 通 "(40))}
```

（b）二つの種文章に対して、それぞれの種文章の概念を包含して、検索用特徴タームと重みを取得した場合

概念検索条件式：

```
{(" 検索条件として入力された文章と概念が似通った文書を検索する。"," 種文章は複数指定できる。")}
```

出力結果：

```
{(" 検索 "(100)," 条件 "(82)," 入力 "(75)," 文章 "(55)," 概念 "(54)," 複数 "(53)," 指定 "(50)," 似 "(43)," 通 "(40)," 文書 "(39))}
```

（c）二つの単純文字列指定（種文章はそれぞれ一つ）に対して、単純文字列指定ごとに検索用特徴タームと重みを取得した場合

概念検索条件式：

```
{(" 検索条件として入力された文章と概念が似通った文書を検索する。"),{" 単純文字列指定は複数指定できる。"}}
```

出力結果：

```
{(" 検索 "(100)," 条件 "(82)," 入力 "(75)," 文章 "(55)," 概念 "(54)," 似 "(43)," 通 "(40),"  
 文書 "(39))},{(" 指定 "(100)," 複数 "(77)," 単純 "(55)," 文字列 "(49))}
```

(8) 注意事項

横分割した表で運用している場合は、「phnchgcon (特徴ターム取得 RD エリア指定ユティリティ)」を使用して、取得する RD エリアを指定してください。指定しない場合には、RD エリアごとにランダムな順番で、検索用特徴タームと重みを返却します。

4.2.11 検索条件式の指定に関する注意

検索条件式として、1文字指定または1文字のワイルドカード指定をした場合は、参照するインデクスが多くなり、システムに負荷を与えることがありますので、ご注意ください。

4.3 全文検索条件式の文法（文書検索プラグインの場合）

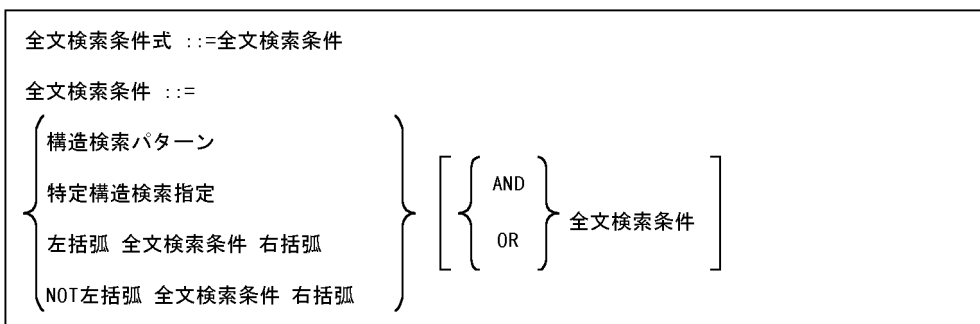
この節では、SGMLTEXT 型のデータ（文書）に対して使用する全文検索条件式の文法について説明します。

4.3.1 構造名を指定した検索条件の AND/OR 論理条件

構造名を指定した検索条件式を構造検索パターンといいます。全文検索条件には、この構造検索パターンの AND/OR 論理条件を指定できます。

全文検索条件式の形式を次の図に示します。

図 4-2 全文検索条件式の形式



注意事項

AND/OR 条件，または NOT 条件を指定する場合，次の優先順位に従って演算します（左側の方が優先順位が高い）。

括弧 > NOT > AND > OR

全文検索条件に AND 条件と OR 条件を同じレベルで記述し，かつ OR 条件を優先したい場合は，OR 条件の範囲を括弧で囲ってください。

指定例

構造名「文章」の下の「概要」に文字列「NETWORK」を含む。

```
文章 [ 概要 { "NETWORK" } ]
```

構造名「文章」の下の「概要」に文字列「NETWORK」を含む文書以外。

```
NOT ( 文章 [ 概要 { "NETWORK" } ] )
```

構造名「文章」の下の「概要」に文字列「COMPUTER」の同義語と「NETWORK」を含む（同義語辞書は「USR01」）。

```
文章 [ 概要 { SYNONYM (USR01, "COMPUTER") AND "NETWORK" } ]
```

構造名「文章」の下の「概要」に文字列「COMPUTER」の同義語を含み，かつ，

4. 抽象データ型関数を使用した文書または文字列データの操作

構造名「文章」の下の「本文」に文字列「NETWORK」または構造名「文章」の下の「コラム」に文字列「DOCUMENT」を含む（同義語辞書は「USR01」、OR条件を優先）。

```
文章 [ 概要 {SYNONYM(USR01, "COMPUTER")} ] AND ( 文章 [ 本文  
{"NETWORK"} ] OR 文章 [ コラム {"DOCUMENT"} ] )
```

4.3.2 構造名の指定

構造検索パターンでは、検索対象とする文書要素をその構造名で指定します。また、構造指定のあとに、その文書要素内での検索条件を指定します。

構造名指定は、検索対象の構造名を文字列で指定します。

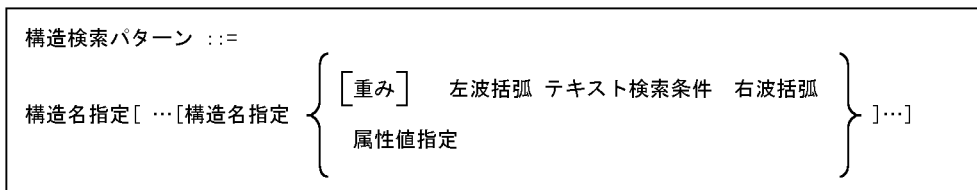
波括弧で囲まれたテキスト検索条件は、指定構造内での全文検索条件を指定する条件式です。

角括弧で構造検索パターンを囲むことで、上位の構造名指定と、指定した上位構造指定以下の構造名をさらに指定した複雑な構造指定ができます。これによって、例えば、要旨の下位構造の目的の部分で、テキスト検索条件を満たす文書を探せます。

属性を持つ構造の属性値を検索条件とする場合は、構造名指定の後ろに属性値の条件を指定します。

構造検索パターンの形式を次の図に示します。

図 4-3 構造検索パターンの形式



注意事項

- 構造名指定は、最上位構造名から順に指定してください。最上位構造名だけを指定した場合、または構造名指定を省略した場合には、文書に含まれるすべての構造を対象に検索します。
- 次のように指定した構造名が未登録の場合はヒットしません。
- 指定した構造を持った文書が登録されていない場合
- 指定した構造名が間違っている場合
- 構造名に半角スペース（0x20）は指定できません。

指定例

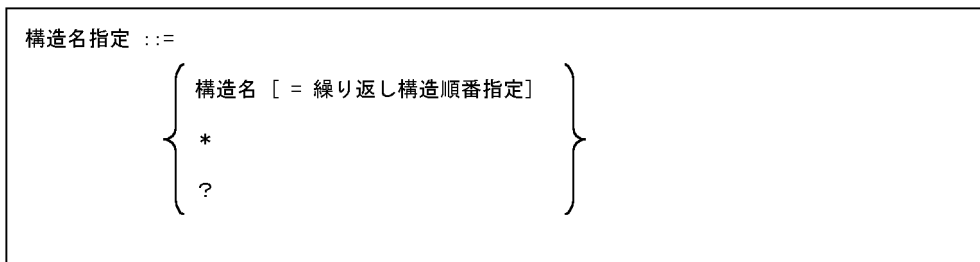
構造名「文章」の下の「概要」の下の「目的」に文字列「NETWORK」を含む。

```
文章 [ 概要 [ 目的 {"NETWORK"} ] ]
```


4.3.3 構造名のワイルドカード指定

構造名は、次の図に示す形式でワイルドカードが指定できます。

図 4-4 構造名指定の形式



* は、0 個以上の任意の構造階層を示します。

? は、1 個の任意の構造階層を示します。

また、構造名指定時に、繰り返して出現する構造名の番号を指定できます。「構造名 =」のあとに、繰り返し構造の順番を指定します。指定方法は次のとおりです。

数 1

数 1 に示された順番を検索対象とします (順番の単独指定)。

数 1-

数 1 以降の順番を検索対象とします (順番の開始指定)。

- 数 1

数 1 までの順番を検索対象とします (順番の最終指定)。

数 1- 数 2

数 1 ~ 数 2 の順番を検索対象とします (順番の範囲指定)。

数 1, 数 2,

指定された数 (数 1, 数 2) の順番を検索対象とします (任意の順番指定)。

注意事項

- "*" (アスタリスク), "?" (疑問符), "=" (イコール), "-" (ハイフン), ",", " (コンマ) および繰り返し構造順番指定の数字は、ASCII コードで指定します。
- 順番指定の範囲は 1 ~ 2,147,483,647 で指定します。
- 順番指定での範囲指定 " - " の値は、数 1- 数 2 の場合、数 1 < 数 2 の関係で指定します。数 1 が数 2 より大きい値の場合はエラーとなります。
- 順番指定での任意の指定 ", " の値に、同じ値を指定した場合、一つの設定とします。
- 例えば、「数 1, 数 2, 数 1」は「数 1, 数 2」となります。

4. 抽象データ型関数を使用した文書または文字列データの操作

- 順番指定での範囲指定 "-" と任意の指定 ", " の混在はできません。エラーとなります。
- 例えば, 「数 1- 数 2, 数 3」はエラーとなります。
- "*" は最下位に指定できません。ただし, すべての階層の属性を指定する場合は除きます。
- 例えば, 「構造名 1[*: 属性名 {" 検索ターム "}]」はエラーとなりますが, 「*: 属性名 {" 検索ターム "}]」はエラーにはなりません。
- "*" 直後の下位構造に "*", "?" は指定できません。
- 例えば, 「*[{ " 検索ターム "}]」のような指定です。

指定例

構造名「文章」の下の「概要」の下の 2 番目の「コラム」に「NETWORK」を含む。

```
文章 [ 概要 [ コラム =2 { "NETWORK" } ] ]
```

構造名「文章」の下の「概要」以下の「コラム」に文字列「NETWORK」を含む。

```
文章 [ 概要 [ * [ コラム { "NETWORK" } ] ] ]
```

構造名「文章」の下の「概要」の一つ下の階層に文字列「NETWORK」を含む。

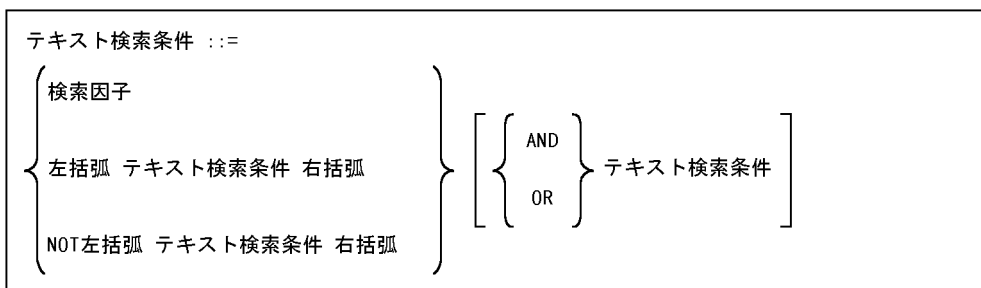
```
文章 [ 概要 [ ? { "NETWORK" } ] ]
```

4.3.4 テキスト検索条件の指定

テキスト検索条件は, 検索因子と呼ぶ条件単位の AND/OR 論理条件によって構成されています。検索因子を使用して, 同一の文書構造内の複雑な検索条件を指定できます。

テキスト検索条件の形式を次の図に示します。

図 4-5 テキスト検索条件の形式



注意事項

AND/OR 条件, または NOT 条件を指定する場合, 次の優先順位に従って演算します (左側の方が優先順位が高い)。

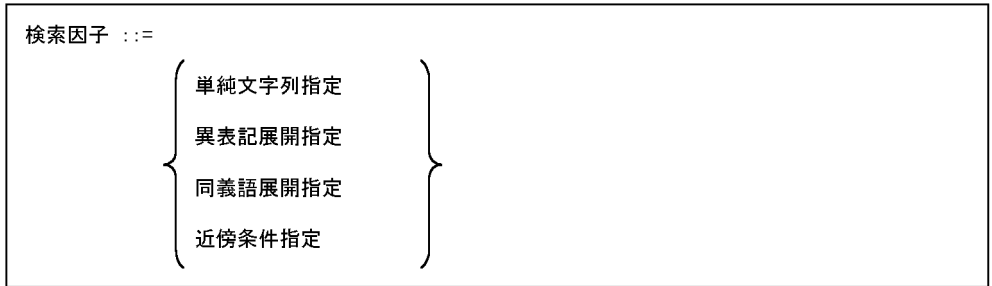
括弧 > NOT > AND > OR

テキスト検索条件に AND 条件と OR 条件を同じレベルで記述し、かつ OR 条件を優先したい場合は、OR 条件の範囲を括弧で囲んでください。

検索因子とは、単純文字列指定、異表記展開指定、同義語展開指定、および近傍条件のどれかの検索条件からなっています。

検索因子の形式を次の図に示します。

図 4-6 検索因子の形式



指定例

文字列「COMPUTER」の同義語または「NETWORK」を含む（同義語辞書は USR01）

```
SYNONYM(USR01, "COMPUTER") OR "NETWORK"
```

文字列「NETWORK」を含まない。

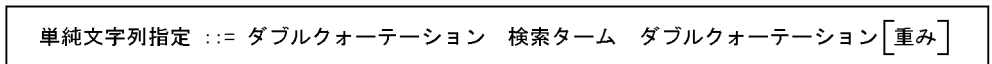
```
NOT ({ "NETWORK" })
```

4.3.5 単純文字列指定

単純文字列指定では、検索したい文字列（検索ターム）をダブルクォーテーションで囲んで指定します。

単純文字列指定の形式を次の図に示します。

図 4-7 単純文字列指定の形式



検索文字列には、ワイルドカードおよび特殊文字を指定できます。ワイルドカードおよび特殊文字は、半角で指定してください。

指定できるワイルドカードおよび特殊文字の意味を次の表に示します。

4. 抽象データ型関数を使用した文書または文字列データの操作

表 4-4 指定できるワイルドカードおよび特殊文字

| 文字 | 文字の意味 |
|----|------------------------|
| * | 0 文字以上の任意の文字列に相当します |
| ? | 任意の 1 文字に相当します |
| | 構造の先頭または末尾に一致します |
| ¥ | ワイルドカードまたは特殊文字の意味を消します |

ワイルドカード指定時の注意事項

ワイルドカードを指定した場合は、同義語展開されません。

指定例

「ネットワーク」と「コンピュータ」で構成される文字列。

" ネットワーク * コンピュータ "

注意事項

同一構造内で出現回数が多い文字を使用したワイルドカード検索または近傍検索を実行した場合、検索時最大ワークメモリサイズを超えてメモリを使用する場合があります。

英文の場合に指定できるワイルドカードおよび特殊文字の意味を次の表に示します。

表 4-5 指定できるワイルドカードおよび特殊文字（英文の場合）

| 文字 | 文字の意味 |
|----|--------------------------|
| * | 単語内の 0 文字以上の任意の文字列に相当します |
| ? | 単語内の任意の 1 文字に相当します |
| ¥ | ワイルドカードまたは特殊文字の意味を消します |

ワイルドカード指定時の注意事項

- ワイルドカードを指定した場合は、同義語展開されません。
- ワイルドカードを使用する場合は、ワイルドカード以外に 3 文字以上指定する必要があります。2 文字以下の場合はエラーになります。

指定例

単語一致検索（「Application」で構成される文字列）

Application

前方一致検索（前方が「app」,「applicatio」で構成される文字列）

app*, applicatio?

また、検索文字列には、除外文字検索用の特殊文字も指定できます。除外文字検索とは、検索タームの一部または前後に特定の文字がある場合を除外して、該当する文書を検索する方法です。除外文字検索用の特殊文字は、半角で指定してください。

ただし、英文検索では除外文字は指定できません。

指定できる除外文字検索用の特殊文字の意味と指定例を次の表に示します。

表 4-6 指定できる除外文字検索用の特殊文字

| 文字 | 文字の意味 | 指定例 |
|----|--|---|
| ^ | 直後にある 1 文字を検索対象から除外します。ただし、特殊文字を "\\\\" (エスケープ文字) とともに指定した 2 文字は 1 文字とみなして除外します。 | <ul style="list-style-type: none"> • 「ローマ?」から「ローマ字」の場合を除外する (? は任意の 1 文字) "ローマ^字" • 「?文書」から「誤文書」の場合を除外する。 "^誤文書" • 「データ??」から「データベース」の場合を除外する。 "データ^ベ^ー^ス" • 「?100」から「¥100」の場合を除外する。 "^¥100" |
| [] | 除外する 1 文字を "," (コンマ) で区切って複数指定する場合に使用します。 | <ul style="list-style-type: none"> • 「大阪?」から「大阪市」と「大阪府」の場合を除外する。 "大阪^[市,府]" |
| , | [] 中の文字を区切る場合に使用します。 | <ul style="list-style-type: none"> • 「文書?管理」から「文書の管理」と「文書を管理」を除外する。 "文書^[の,を]管理" |

除外文字検索用の特殊文字指定時の注意事項

- 検索ターム全体または "*" で区切られた部分に除外する文字以外の文字がない場合は、エラーとなります。
誤った例: "^ア", "[ア,イ]", "ア*", "ア*イ", "ア*^イ", "?^ア"
正しい例: "ア^イ", "[ア,イ]ウ", "イ^ア*", "ウ^ア*イ",
"ア*^イウ", "?^アイ"
- 除外文字検索用の特殊文字を指定した場合、同義語展開されません。

英文検索の場合の注意事項

- 検索文字列は 1 バイト文字コードの '0x20' ~ '0x7E' の範囲で指定します。それ以外の 1 バイト文字コード、または 2 バイト文字コードを指定した場合、そのコードはないものとして扱い、以降の文字コードを詰めて検索します。指定した文字コードが、英文検索用インデックス作成時に定義した不要語、または削除コードと同じコードの場合は、エラーとなります。
- 検索文字列にインデクスデリミタを含む場合は、インデクスデリミタで前後の文字コードを区切り、フレーズ検索とみなして検索します。例えば、'0x20' (半角スペース) で区切って複数の単語を記述した場合、それらの単語が隣り合って出現する一つの単語列として検索します。なお、文字列中の一単語は、64 文字 (バイト) まで指定できます。64 文字 (バイト) を超える単語が指定された場合は、65 文字 (バイト) 目以降を無視して、64 文字 (バイト) 目までを単語として検索します。

4.3.6 異表記展開指定

単純文字列指定では、例えば、検索対象文字列に「ボウリング」を指定した場合、「ボウリング」だけを検索対象とします。これに対し、異表記展開指定では、検索対象文字列を同義語異表記の文字列に展開し、その文字列も検索対象とします。例えば、「ボウリング」を指定した場合、「ボウリング」だけではなく、「ボーリング」や「ボウリング」などを検索することができます。英文検索の異表記展開は、半角アルファベット文字列を対象とした、半角アルファベット大小文字列展開をして文字列を検索します。このように、異表記展開を指定すると、より広い範囲での文字列検索が可能となります。

異表記展開指定の形式を次の図に示します。

図 4-8 異表記展開指定の形式

異表記展開指定 ::=

SOUNDEX_EXP 左括弧 単純文字列指定 [, 展開モード指定] 右括弧

展開モードは、展開する異表記に応じて指定します。異表記オプションの種類と指定する展開モードの対応を次の表に示します。

表 4-7 異表記オプション

| 異表記オプション | 展開モード | 説明 |
|--------------|-------|--|
| アルファベット異表記 | A | 全角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |
| かたかな異表記 | K | 全角かたかな文字列に対して、かたかな異表記の展開をします。 |
| かたかな長音ハイフン展開 | L | かたかな文字列に対して、長音・ハイフン展開をします。このとき、展開されるのは長音・ハイフンだけであり、前後のかたかな文字列は展開の対象となりません。 |
| かたかなベーシック展開 | B | かたかなルールに基づいた展開だけを実行します。 |
| かたかな拗音展開 | C | かたかな文字列に対して、拗音展開をします。 |
| 全角半角異表記 | E | 指定した文字列中で全角から半角にできるもの、または半角から全角にできるものに対して、次の 3 パターンに展開します。 1. すべてが全角文字列 2. すべてが半角文字列 3. 指定した文字列 |

| 異表記オプション | 展開モード | 説明 |
|-------------|-------|--|
| 半角アルファベット展開 | H | 半角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |
| ひらがな拗音展開 | Y | ひらがなの拗音の展開をします。 |
| 英単語派生表記展開 | S | 半角英文字の文字列を派生語に展開します。 |
| 英単語語尾表記展開 | I | 英単語を語尾の変換系に展開します。 |

注 このオプションを使用するには、Text Search Enhancer for English が必要です。

これら展開モードは単独で指定するだけでなく、組み合わせで指定することもできます。例えば、全角文字のアルファベット大小文字展開、および全角半角文字展開をする場合は、"AE" と指定します。展開モードの指定がない場合は、"AKE" が指定されたと仮定されます。なお、"BCL" を指定した場合は、"K" を指定したときと同じ展開となります。

展開モードは、順不同で指定できます。必要な組み合わせを指定してください。

異表記展開指定時の注意事項

長大な英文に対してアルファベット大小文字展開を実行したり、長大なかな文字列に対してかな異表記展開を実行したりすると、展開数の増加によって、多大なリソースが必要になったり、レスポンスが低下したりすることがあります。

異表記展開指定をする場合は、次のような点に留意してください。

- エンドユーザが自由に検索条件を指定できるようなシステムの場合、環境設定時に、環境定義ファイルで同義語異表記展開用最大メモリサイズを設定しておくことをお勧めします。これによって、異表記展開数が多大になった場合にエラーで中断させることができ、システムのリソース消費量を制限できます。
- 英文に対して、アルファベット大小文字展開を必ず実行するような運用の場合は、異表記展開指定ではなく、大小文字統一機能を使用することをお勧めします。大小文字統一機能は、インデクス定義時に定義します。

英文に対して大小文字展開を指定すると、それぞれの単語に対して大小文字展開が実行され、そのすべての組み合わせの OR 検索が実行されます。展開数は、「1 単語当たり三つの展開形式 + 指定文字列」の 4 パターンとなり、「4 の n 乗 (n は単語数)」になります。ただし、指定された文字が展開パターンと一致する場合は 3 パターンとなり、展開数は「3 の n 乗 (n は単語数)」になります。

展開例を次に示します。

指定した英文

long long ago ,there were three little pigs.

アルファベット大小文字展開の結果

long long ago ,there were three little pigs.

4. 抽象データ型関数を使用した文書または文字列データの操作

```
Long long ago ,there were three little pigs.  
LONG long ago ,there were three little pigs.  
long Long ago ,there were three little pigs.  
:
```

この場合の展開数は、「展開数 = $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 = 3$ の 8 乗 = 6,561 パターン」となります。さらに、全角半角異表記展開を同時に指定した場合は、それぞれの展開結果をすべて全角とすべて半角の 2 パターンに展開するため、展開数が 2 倍になります。

展開数の増加は、リソースの増加およびレスポンスの低下につながるため、このような検索は避けてください。

- 長大なかたかな文字列に対して、かたかな異表記展開を必ず実行するような運用の場合は、拗音統一機能を使用してかたかな展開数を削減するようにしてください。拗音統一機能は、インデクス定義時に定義します。また、環境設定時に、環境定義ファイルで同義語異表記展開用最大メモリサイズを設定して、リソース消費量の上限值を設定してください。

長大なかたかな文字列に対してかたかな異表記展開を指定すると、かたかな文字列のそれぞれの部分に対してかたかな異表記展開が実行され、そのすべての組み合わせの OR 検索が実行されます。なお、かたかな異表記展開は、アルファベット大小文字展開や全角半角展開のように明確な規則に従った展開でなく、TextSearch Plug-in で設定されているルールに基づいて展開されます。このため、かたかな文字列によって展開数は異なります。

展開例を次に示します。

指定したかたかな文字列

バイクバイクバイクバイクバイク

かたかな異表記展開の結果

「バイク」は「ヴァイク」、「ヴァイク」、「ヴァイク」、「ヴァイク」、「バイク」、「バイク」の 6 パターンに展開されるため、それぞれ「バイク」の部分を 6 パターンに展開した組み合わせが検索条件になります。したがって、例のように「バイク」が 5 回つながっているようなかたかな文字列の展開数は、「展開数 = $6 \times 6 \times 6 \times 6 \times 6 = 7,776$ パターン」になります。

展開数の増加は、リソースの増加およびレスポンスの低下につながるため、このような検索は避けてください。

- 展開速度を落とさずに検索を実行するには、部分展開検索モードを指定して検索を実行してください。この検索モードを使用するかどうかは、環境定義ファイルで指定します。指定方法については、「2.6.2(2) 設定できる項目」を参照してください。なお、部分展開検索モードで検索を実行すると、システムのリソース消費量が増加します。このため、検索最大ワークメモリサイズ (phn_mem_search) の設定値を見直してください。検索最大ワークメモリサイズで指定したメモリサイズを超えた場合には、メモリ不足エラーとするか、または検索用ワークファイルを使用するかを指定できます。

指定例

- 文字列「COMPUTER」のアルファベット大小文字および全角半角文字展開を含

む。

```
SOUNDEX_EXP("COMPUTER", "AE")
```

異表記展開の例を次に示します。

| 指定モード | 検索文字列 | 展開後文字列 |
|-------|-----------|--|
| A | MAX | MAX, max, Max |
| | HiRDB | HIRDB, hirddb, Hirddb, HiRDB |
| B | フェーズ | フェーズ, フェイズ |
| C | フェーズ | フェーズ, フェーズ |
| E | Group | Group, Group |
| | テレビ会議システム | テレビ会議システム, テレビ会議システム |
| | Uターン | Uターン, Uターン |
| H | MAX | MAX, max, Max |
| | HiRDB | HIRDB, hirddb, Hirddb, HiRDB |
| I | sing | sings, sang, sung, singing |
| K | バイク | ヴァイク, ヴァイク, ヴァイク, ヴァイク, バイク, バイク |
| L | フェーズ | フェーズ, フェーズ |
| S | absorb | absorption, absorbed, absorbing, absorbedly, absorbinglyなど |
| Y | ちょう | ちょう, ちょう, ちょう, ちょう |
| AE | MAX | MAX, max, Max, MAX, max, Max |
| | Uターン | Uターン, uターン, Uターン, uターン |
| | Uターン | Uターン, uターン, Uターン, uターン, Uターン, uターン |
| KE | バイク | ヴァイク, ヴァイク, ヴァイク, ヴァイク, バイク, バイク, ヴァイク, ヴァイク, ヴァイク, ヴァイク, バイク, バイク |
| | Uターン | Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン |
| | Uターン | Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン |
| HE | MAX | MAX, max, Max, MAX, max, Max |
| | Uターン | Uターン, uターン, Uターン, uターン |
| | Uターン | Uターン, uターン, Uターン, uターン, Uターン |
| SI | sing | singing, singing's, singable, sing, sings, sang, sung, singer, singer's, singers, singers' |
| AKE | Uターン | Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, uターン, uターン |
| | Uターン | Uターン, Uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, Uターン, Uターン, uターン, uターン, Uターン, Uターン, uターン, uターン |

4.3.7 同義語展開指定

同義語展開指定では、同義語展開に使用する辞書名と検索タームを指定します。また、異表記展開モードの指定もできます。異表記展開モードを指定した場合は、同義語展開したすべての語に関して、指定のモードで異表記展開します。異表記展開モードの指定

4. 抽象データ型関数を使用した文書または文字列データの操作

がない場合は、同義語展開の結果で検索します。

同義語展開指定の形式を次の図に示します。

図 4-9 同義語展開指定の形式

| |
|--|
| 同義語展開指定 ::= SYNONYM 左括弧 同義語辞書名, 単純文字列指定 [, 展開モード指定] 右括弧 |
|--|

同義語展開指定時の注意事項

- 検索文字列にワイルドカード除外文字検索用の特殊文字を指定した場合は、同義語展開されません。

指定例

- 文字列「COMPUTER」の同義語、アルファベット大小文字および全角半角展開を含む（同義語辞書は USR01）。

```
SYNONYM( USR01, "COMPUTER", "AE" )
```

4.3.8 近傍条件指定

近傍条件指定では、二つの検索ターム（語句リスト）とそれらの距離条件、出現順序関係を指定します。それぞれの語句リスト中の検索タームのどれかがヒットした位置で、距離条件に当てはまるように近接している語句リストがヒットする文書を検索します。

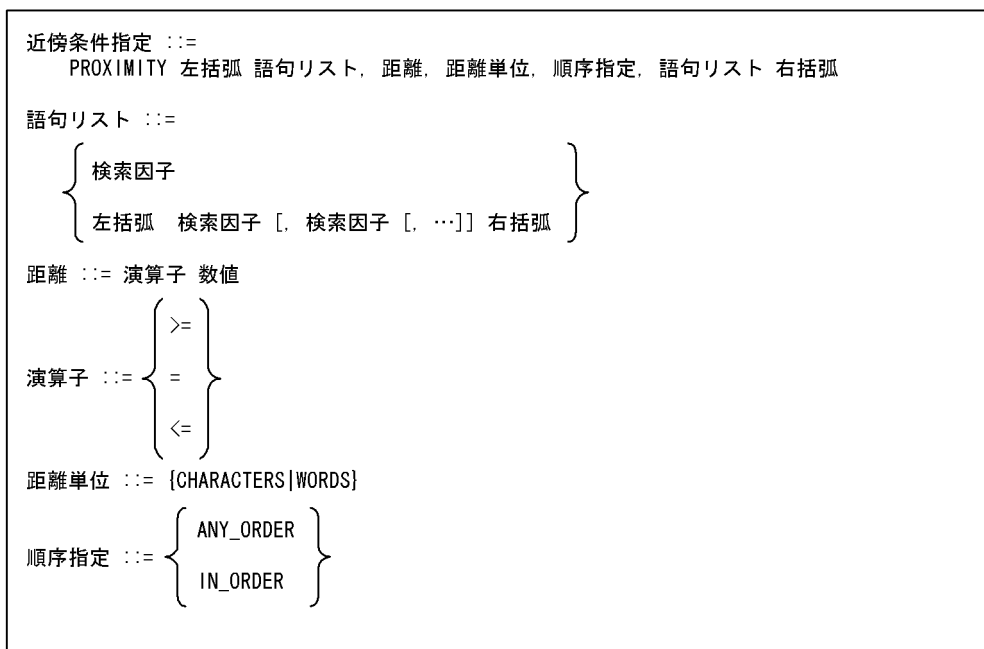
距離条件として、距離には演算子と数値、距離単位には文字を表す 'CHARACTERS' を使用します。これによって、二つの検索ターム間の文字数（距離）を検索条件にできます。

演算子は、指定距離に等しい・以上・以下の3種類があり、等しい場合は = を、以上の場合は >= を、以下の場合は <= を指定します。

順序指定は、第一の語句リストと第二の語句リストのどちらが先に出現してもよいとする 'ANY_ORDER' と、第一の語句リストの次に第二の語句リストが出現する順序を指定する 'IN_ORDER' を選択できます。

近傍条件指定の形式を次の図に示します。

図 4-10 近傍条件指定の形式



指定例

文字列「COMPUTER」の同義語，アルファベット大小文字および全角半角文字展開と，文字列「ネットワーク」の同義語，アルファベット大小文字，かな異表記，全角半角文字展開が 20 文字以下に近接して出現する（同義語辞書は USR01）。

```

PROXIMITY(SYNONYM( USR01, "COMPUTER", "AE"), <=20,
CHARACTERS, ANY_ORDER, SYNONYM( USR01, "ネットワーク", "AKE"))

```

英文検索用インデックスの場合

英文検索用インデックスは距離条件として，距離には演算子と数値，距離単位には単語を表す 'WORDS' を使用します。

英文検索用インデックスに対して距離単位に 'CHARACTERS' を指定した場合はエラーとなります。

近傍演算の距離は 0 ~ 2,147,483,646 の範囲で指定します。距離の単位は単語です。インデックスデリミタで区切られた単語を 1 単語とみなします。例えば，英文書に "I have a pen." とある場合，インデックスデリミタが半角スペースのため，'I' と 'pen' の距離は 2 となります。また，不要語が含まれる場合，不要語も 1 単語として近傍演算の距離に数えます。

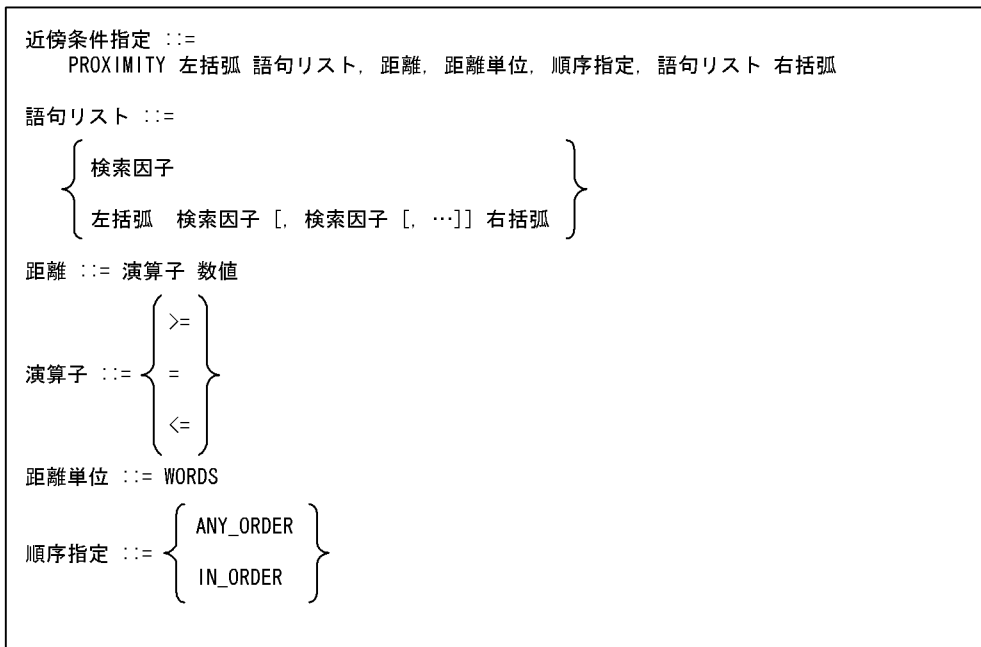
演算子は，指定距離に等しい・以上・以下の 3 種類があり，等しい場合は = を，以上の場合は >= を，以下の場合は <= を指定します。

順序指定は，第一の語句リストと第二の語句リストのどちらが先に出現してもよいとする 'ANY_ORDER' と，第一の語句リストの次に第二の語句リストが出現する順序を指定する 'IN_ORDER' を選択できます。

4. 抽象データ型関数を使用した文書または文字列データの操作

近傍条件指定の形式を次の図に示します。

図 4-11 英文インデックスの近傍条件指定の形式



指定例

文字列「COMPUTER」と「SYSTEM」が10単語以下に近接して出現する。

```
PROXIMITY("COMPUTER", <= 10, WORDS, ANY_ORDER, "SYSTEM")
```

4.3.9 重み

重みは、複数の検索タームを指定する場合、スコア値を算出するときに、より重要度を高くしたい方の検索タームに付けます。これによって、重みを付けた検索タームにはより高いスコアが付き、重みを指定できる範囲は、1 ~ 100 です。

重み指定の形式を次の図に示します。

図 4-12 重み指定の形式



なお、重みは、contains_with_score 関数の全文検索条件式にだけ指定できます。contains 関数の全文検索条件式および extracts 関数のハイライト位置条件式で重みを指定しても無視されます。contains_with_score 関数、contains 関数、および extracts 関数の詳細については、「4.2 抽象データ型関数の文法」を参照してください。

指定例

- 構造名「文章」の下の「本文」に文字列「コンピュータ」を含み、かつ構造名「文章」の下の「概要」に文字列「ネットワーク」を含む。スコア値を算出する際は、文字列「コンピュータ」の出現数に 10 を掛ける。

文章[本文(10){"コンピュータ"}] AND文章[概要{"ネットワーク"}]

- 構造名「文章」の下の「本文」に文字列「コンピュータ」または「ネットワーク」を含む。スコア値を算出する際は、文字列「ネットワーク」の出現数に 10 を掛ける。

文章[本文{"コンピュータ" OR "ネットワーク"}(10)]

4.3.10 属性値指定

構造に指定された属性を条件にして検索する場合、構造に対する属性名およびテキスト検索条件を指定します。

属性値指定の形式を次の図に示します。

図 4-13 属性値指定の形式

属性値指定 ::= コロン 属性名指定 [重み] 左波括弧 テキスト検索条件 右波括弧

指定例

構造名「文章」の属性「Author」に「Tanaka」を含む。

文章:Author{"Tanaka"}

4.3.11 特定構造検索指定

特定構造検索指定は、複数の構造を対象とした検索条件を満たす構造が、ある特定構造に含まれているかどうかを検索する場合に使用します。

特定構造検索指定では、検索の対象となる特定構造および検索条件式を二つ以上指定します。ただし、特定構造検索指定の構造検索パターンは、属性値指定できません。

特定構造検索指定の形式を次の図に示します。

図 4-14 特定構造検索指定の形式

特定構造検索指定 ::=

SPECIFIC_AND 左括弧 特定構造指定: 構造検索パターン, 構造検索パターン
[, 構造検索パターン[, …]] 右括弧

特定構造指定には、検索範囲を特定するための構造名を指定します。

特定構造指定の形式を次の図に示します。

4. 抽象データ型関数を使用した文書または文字列データの操作

図 4-15 特定構造指定の形式

```
特定構造指定 ::=  
構造名指定 [...[構造名指定]...]
```

指定例

最上位構造「報告書」の下位構造「会議」の下位に、「所属」と「出席者」という構造を持つ文書について、「所属」に文字列「横浜営業所」を含み、かつ「出席者」に文字列「鈴木」を含む文書を検索する。

```
SPECIFIC_AND(報告書[会議]; 報告書[会議[所属{"横浜営業所"}]], 報告書[会議[出席者{"鈴木"}]])
```

4.3.12 特殊文字の指定

全文検索条件の条件式中で特殊な意味を持つ記号（特殊記号）は、単独では文字として指定できません。文字として指定する場合は、各特殊記号の前に"¥"（エスケープ文字）を付けてください。

全文検索条件の単純文字列指定以外と、単純文字列指定内では条件式中の特殊記号が異なります。

"¥"（エスケープ文字）を付ける必要のある特殊記号を表 4-8 および表 4-9 に示します。

表 4-8 全文検索条件の条件式中に指定できる特殊記号（全文検索条件の単純文字列指定以外）

| 特殊記号 | 文字の呼称 |
|------|--------------|
| " | ダブルクォーテーション |
| | スペース |
| , | コンマ |
| | タブ |
| { | 左波括弧 |
| } | 右波括弧 |
| [| 左角括弧 |
|] | 右角括弧 |
| (| 左丸括弧 |
|) | 右丸括弧 |
| ¥ | 円記号（エスケープ文字） |
| : | コロン |
| * | アスタリスク |
| ? | 疑問符 |

| 特殊記号 | 文字の呼称 |
|------|-------|
| = | イコール |
| - | ハイフン |

表 4-9 全文検索条件の条件式中に指定できる特殊記号（全文検索条件の単純文字列指定内）

| 特殊記号 | 文字の呼称 |
|------|--------------|
| * | アスタリスク |
| ? | 疑問符 |
| | パイプ |
| ¥ | 円記号（エスケープ文字） |
| [| 左角括弧 |
|] | 右角括弧 |
| , | コンマ |
| (| 左丸括弧 |
|) | 右丸括弧 |
| " | ダブルクォーテーション |
| ^ | ハット |

4.3.13 全文検索条件式の指定例

全文検索条件式の指定例を次に示します。なお、例では表「reports」の SGMLTEXT 型の列「doc」を検索対象としています。また、「概要」および「目的」は、最上位構造「文章」の下位構造です。

「概要」部分に文字列「COMPUTER」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(doc, '文章[概要{"COMPUTER"}]') IS TRUE
```

「概要」部分に文字列「COMPUTER」の同義語を含む文書のタイトル一覧表を作成する。

```
SELECT title FROM reports
WHERE contains(doc, '文章[概要{SYNONYM(USR01, "COMPUTER")}]') IS TRUE
```

「概要」部分に文字列「COMPUTER」の同義語と「NETWORK」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(doc, '文章[概要{SYNONYM(USR01, "COMPUTER")
AND "NETWORK"}]') IS TRUE
```

4. 抽象データ型関数を使用した文書または文字列データの操作

「概要」部分に「COMPUTER」の同義語と「NETWORK」の同義語を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
  WHERE contains(doc, '文章[概要{SYNONYM(USR01,
    "COMPUTER")} AND SYNONYM(USR01,
    "NETWORK")}]' ) IS TRUE
```

「概要」部分に「COMPUTER」の同義語と「NETWORK」の同義語が10文字以内に近接して出現する文書のタイトル一覧表を作成する。

```
SELECT title FROM reports
  WHERE contains(doc, '文章[概要{PROXIMITY(SYNONYM
    (USR01, "COMPUTER"), <=10, CHARACTERS,
    ANY_ORDER, SYNONYM(USR01,
    "NETWORK"))}]' ) IS TRUE
```

「概要」部分に、「COMPUTER」の同義語（同義語辞書USR01）と、「NETWORK」の同義語（同義語辞書POLITICS）が20文字以内に近接して出現する文書のタイトル、著者一覧表を作成する。

```
SELECT title, author FROM reports
  WHERE contains(doc, '文章[概要{PROXIMITY(SYNONYM
    (USR01, "COMPUTER"), <= 20, CHARACTERS,
    ANY_ORDER, SYNONYM( POLITICS,
    "NETWORK"))}]' ) IS TRUE
```

「目的」部分に「COMPUTER」を含み、かつ「概要」部分に「NETWORK」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
  WHERE contains(doc, '文章[目的{"COMPUTER"}] AND文章[概要
    {"NETWORK"}]' ) IS TRUE
```

「文章」に「NETWORK」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
  WHERE contains(doc, '文章{"NETWORK"}' ) IS TRUE
```

「文章」の下の任意の階層にある「コラム」部分に文字列「COMPUTER」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
  WHERE contains(doc, '文章[*[コラム{"COMPUTER"}]]' ) IS TRUE
```

3階層目にある「コラム」部分に文字列「COMPUTER」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
  WHERE contains(doc, '?[?[コラム{"COMPUTER"}]]' ) IS TRUE
```

「文章」の下階層にある2番目の「コラム」部分に文字列「COMPUTER」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
```


4. 抽象データ型関数を使用した文書または文字列データの操作

```
WHERE contains(doc, '文章[コラム=2{"COMPUTER"}]') IS TRUE
```

「文章」に「COMPUTER」と「NETWORK」のアルファベット大小文字展開と全角半角展開を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(doc, '文章{SOUNDEX_EXP("COMPUTER","AE") AND
                    SOUNDEX_EXP("NETWORK","AE")}') IS TRUE
```

4.4 全文検索条件式の文法（文字列検索プラグインの場合）

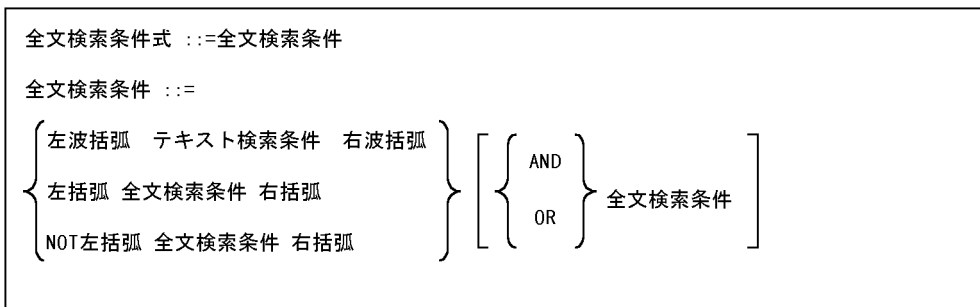
この節では、FREEWORD 型のデータ（文字列データ）に対して使用する全文検索条件式の文法について説明します。

4.4.1 検索条件の AND/OR 論理条件

全文検索条件には、AND/OR 論理条件を指定できます。

全文検索条件式の形式を次の図に示します。

図 4-16 全文検索条件式の形式



注意事項

- AND/OR 条件，または NOT 条件を指定する場合，次の優先順位に従って演算します（左側の方が優先順位が高い）。
括弧 > NOT > AND > OR
- 全文検索条件に AND 条件と OR 条件を同じレベルで記述し，かつ OR 条件を優先したい場合は，OR 条件の範囲を括弧で囲んでください。

指定例

文字列「NETWORK」を含む。

```
{"NETWORK"}
```

文字列「NETWORK」を含む文書以外。

```
NOT ({ "NETWORK" })
```

文字列「COMPUTER」の同義語を含み，かつ「NETWORK」または「DOCUMENT」を含む（同義語辞書は「USR01」，OR 条件を優先）。

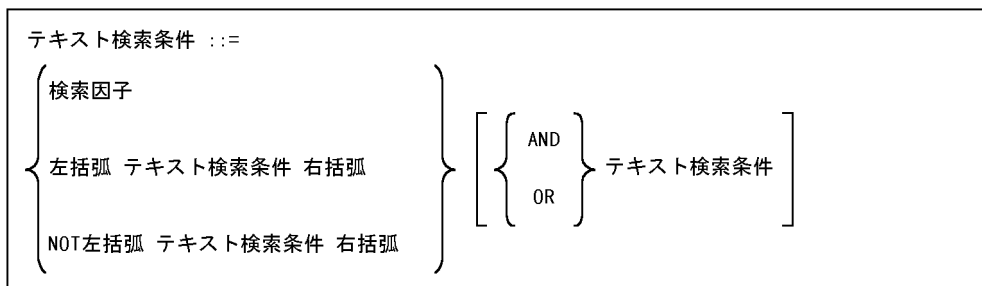
```
{SYNONYM(USR01, "COMPUTER")} AND ({"NETWORK"} OR {"DOCUMENT"})
```

4.4.2 テキスト検索条件の指定

テキスト検索条件は、検索因子と呼ぶ条件単位の AND/OR 論理条件によって構成されています。検索因子を使用して、同一の文書構造内の複雑な検索条件を指定できます。

テキスト検索条件の形式を次の図に示します。

図 4-17 テキスト検索条件の形式



注意事項

AND/OR 条件、または NOT 条件を指定する場合、各演算子について次の優先順位に従って演算します（左側の方が優先順位が高い）。

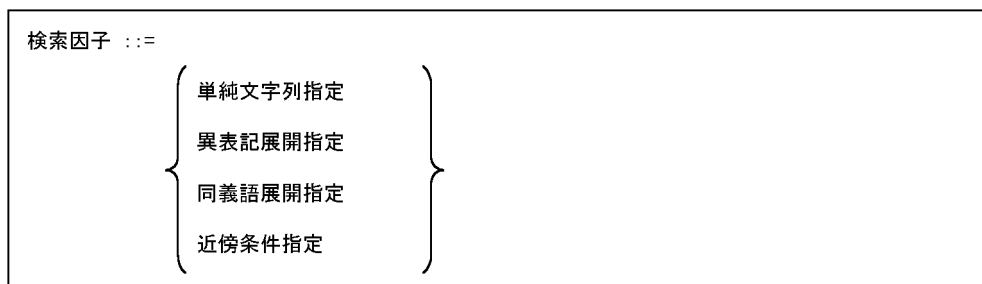
括弧 > NOT > AND > OR

テキスト検索条件に AND 条件と OR 条件を同じレベルで記述し、かつ OR 条件を優先したい場合は、OR 条件の範囲を括弧で囲んでください。

検索因子とは、単純文字列指定、異表記展開指定、同義語展開指定、および近傍条件のどれかの検索条件から成っています。

検索因子の形式を次の図に示します。

図 4-18 検索因子の形式



指定例

文字列「COMPUTER」の同義語または「NETWORK」を含む（同義語辞書は USR01）

SYNONYM(USR01, "COMPUTER") OR "NETWORK"

4. 抽象データ型関数を使用した文書または文字列データの操作

文字列「NETWORK」を含まない。

```
NOT ({ "NETWORK" })
```

4.4.3 単純文字列指定

単純文字列指定では、検索したい文字列（検索ターム）をダブルクォーテーションで囲んで指定します。

単純文字列指定の形式を次の図に示します。

図 4-19 単純文字列指定の形式

| |
|--|
| 単純文字列指定 ::= ダブルクォーテーション 検索ターム ダブルクォーテーション [重み] |
|--|

検索文字列には、ワイルドカードおよび特殊文字を指定できます。ワイルドカードおよび特殊文字は、半角で指定してください。

指定できるワイルドカードおよび特殊文字については、「4.3.5 単純文字列指定」の表 4-4 を参照してください。

また、検索文字列には、除外文字検索用の特殊文字も指定できます。除外文字検索とは、検索タームの一部または前後に特定の文字がある場合を除外して、該当する文書を検索する方法です。除外文字検索用の特殊文字は、半角で指定してください。

ただし、英文検索では除外文字は指定できません。

指定できる除外文字検索用の特殊文字については、「4.3.5 単純文字列指定」の表 4-6 を参照してください。

英文検索の場合の注意事項

- 検索文字列は 1 バイト文字コードの '0x20' ~ '0x7E' の範囲で指定します。それ以外の 1 バイト文字コード、または 2 バイト文字コードを指定した場合、そのコードはないものとして扱い、以降の文字コードを詰めて検索します。指定した文字コードが、英文検索用インデクス作成時に定義した不要語、または削除コードと同じコードの場合は、エラーとなります。
- 検索文字列にインデクスデリミタを含む場合は、インデクスデリミタで前後の文字コードを区切り、フレーズ検索とみなして検索します。例えば、'0x20'（半角スペース）で区切って複数の単語を記述した場合、それらの単語が隣り合って出現する一つの単語列として検索します。なお、文字列中の一単語は、64 文字（バイト）まで指定できます。64 文字（バイト）を超える単語が指定された場合は、65 文字（バイト）目以降を無視して、64 文字（バイト）目までを単語として検索します。

4.4.4 異表記展開指定

単純文字列指定では、例えば、検索対象文字列に「ボウリング」を指定した場合、「ボウリング」だけを検索対象とします。これに対し、異表記展開指定では、検索対象文字列を同義語異表記の文字列に展開し、その文字列も検索対象とします。例えば、「ボウリング」を指定した場合、「ボウリング」だけではなく、「ボーリング」や「ポウリング」などを検索することができます。英文検索の異表記展開は、半角アルファベット文字列を対象とした、半角アルファベット大小文字列展開をして文字列を検索します。このように、異表記展開を指定すると、より広い範囲での文字列検索が可能となります。

異表記展開指定の形式を次の図に示します。

図 4-20 異表記展開指定の形式

異表記展開指定 ::=

SOUNDEX_EXP 左括弧 単純文字列指定 [, 展開モード指定] 右括弧

展開モードは、展開する異表記に応じて指定します。異表記オプションの種類と指定する展開モードの対応を次の表に示します。

表 4-10 異表記オプション

| 異表記オプション | 展開モード | 説明 |
|--------------|-------|--|
| アルファベット異表記 | A | 全角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |
| かたかな異表記 | K | 全角かたかな文字列に対して、かたかな異表記の展開をします。 |
| かたかな長音ハイフン展開 | L | かたかな文字列に対して、長音・ハイフン展開をします。このとき、展開されるのは長音・ハイフンだけであり、前後のかたかな文字列は展開の対象となりません。 |
| かたかなベーシック展開 | B | かたかなルールに基づいた展開だけを実行します。 |
| かたかな拗音展開 | C | かたかな文字列に対して、拗音展開をします。 |
| 全角半角異表記 | E | 指定した文字列中で全角から半角にできるもの、または半角から全角にできるものに対して、次の 3 パターンに展開します。 1. すべてが全角文字列 2. すべてが半角文字列 3. 指定した文字列 |
| 半角アルファベット展開 | H | 半角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |

4. 抽象データ型関数を使用した文書または文字列データの操作

| 異表記オプション | 展開モード | 説明 |
|-----------|-------|----------------------|
| ひらがな拗音展開 | Y | ひらがなの拗音の展開をします。 |
| 英単語派生表記展開 | S | 半角英文字の文字列を派生語に展開します。 |
| 英単語語尾表記展開 | I | 英単語を語尾の変換系に展開します。 |

注 このオプションを使用するには、Text Search Enhancer for English が必要です。

異表記オプションの種類と指定する展開モードの対応の詳細については、表 4-7 を参照してください。

展開モードは単独で指定するだけでなく、組み合わせで指定することもできます。例えば、全角文字のアルファベット大小文字展開、および全角半角文字展開をする場合は、「AE」と指定します。展開モードの指定がない場合は、「AKE」が指定されていると仮定されます。なお、「BCL」を指定した場合は、「K」を指定したときと同じ展開となります。

展開モードは、順不同で指定できます。必要な組み合わせを指定してください。

なお、異表記展開の例については、「4.3.6 異表記展開指定」を参照してください。

異表記展開指定時の注意事項

長大な英文に対してアルファベット大小文字展開を実行したり、長大なかな文字列に対してかな異表記展開を実行したりすると、展開数の増加によって、多大なリソースが必要になったり、レスポンスが低下したりすることがあります。

異表記展開指定をする場合は、次のような点に留意してください。

- エンドユーザが自由に検索条件を指定できるようなシステムの場合、環境設定時に、環境定義ファイルで同義語異表記展開用最大メモリサイズを設定しておくことをお勧めします。これによって、異表記展開数が多大になった場合にエラーで中断させることができ、システムのリソース消費量を制限できます。
- 英文に対して、アルファベット大小文字展開を必ず実行するような運用の場合は、異表記展開指定ではなく、大小文字統一機能を使用することをお勧めします。大小文字統一機能は、インデクス定義時に定義します。

英文に対して大小文字展開を指定すると、それぞれの単語に対して大小文字展開が実行され、そのすべての組み合わせの OR 検索が実行されます。展開数は、「1 単語当たり三つの展開形式 + 指定文字列」の 4 パターンとなり、「4 の n 乗 (n は単語数)」になります。ただし、指定された文字が展開パターンと一致する場合は 3 パターンとなり、展開数は「3 の n 乗 (n は単語数)」になります。

展開例を次に示します。

指定した英文

long long ago ,there were three little pigs.

アルファベット大小文字展開の結果

long long ago ,there were three little pigs.
 Long long ago ,there were three little pigs.
 LONG long ago ,there were three little pigs.
 long Long ago ,there were three little pigs.

:

この場合の展開数は、「展開数 = $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 = 3$ の 8 乗 = 6,561 パターン」となります。さらに、全角半角異表記展開を同時に指定した場合は、それぞれの展開結果をすべて全角とすべて半角の 2 パターンに展開するため、展開数が 2 倍になります。

展開数の増加は、リソースの増加およびレスポンスの低下につながるもので、このような検索は避けてください。

- 長大なかたかな文字列に対して、かたかな異表記展開を必ず実行するような運用の場合は、拗音統一機能を使用してかたかな展開数を削減するようにしてください。拗音統一機能は、インデックス定義時に定義します。また、環境設定時に、環境定義ファイルで同義語異表記展開用最大メモリサイズを設定して、リソース消費量の上限値を設定してください。

長大なかたかな文字列に対してかたかな異表記展開を指定すると、かたかな文字列のそれぞれの部分に対してかたかな異表記展開が実行され、そのすべての組み合わせの OR 検索が実行されます。なお、かたかな異表記展開は、アルファベット大小文字展開や全角半角展開のように明確な規則に従った展開でなく、TextSearch Plug-in で設定されているルールに基づいて展開されます。このため、かたかな文字列によって展開数は異なります。

展開例を次に示します。

指定したかたかな文字列

バイクバイクバイクバイクバイク

かたかな異表記展開の結果

「バイク」は「ヴァイク」、「ヴァイク」、「ヴァイク」、「ヴァイク」、「バイク」、
「バイク」の 6 パターンに展開されるため、それぞれ「バイク」の部分 6 パターンに展開した組み合わせが検索条件になります。したがって、例のように「バイク」が 5 回つながっているようなかたかな文字列の展開数は、「展開数 = $6 \times 6 \times 6 \times 6 \times 6 = 7,776$ パターン」になります。

展開数の増加は、リソースの増加およびレスポンスの低下につながるもので、このような検索は避けてください。

- 展開速度を落とさずに検索を実行するには、部分展開検索モードを指定して検索を実行してください。この検索モードを使用するかどうかは、環境定義ファイルで指定します。指定方法については、「2.8.1(2) 設定できる項目」を参照してください。なお、部分展開検索モードで検索を実行すると、システムのリソース消費量が増加します。このため、検索最大ワークメモリサイズ (phx_mem_search) の設定値を見直してください。

指定例

文字列「COMPUTER」のアルファベット大小文字および全角半角文字展開を含む。

```
SOUNDEX_EXP("COMPUTER", "AE" )
```

```
SOUNDEX_EXP("COMPUTER", "SH")
```

英単語派生表記展開、英単語語尾表記展開の詳細については、マニュアル「Text

4. 抽象データ型関数を使用した文書または文字列データの操作

Search Enhancer for English」を参照してください。

4.4.5 同義語展開指定

同義語展開指定では、同義語展開に使用する辞書名と検索タームを指定します。また、異表記展開モードの指定もできます。異表記展開モードを指定した場合は、同義語展開したすべての語に関して、指定のモードで異表記展開します。異表記展開モードの指定がない場合は、同義語展開の結果で検索します。

同義語展開指定の形式を次の図に示します。

図 4-21 同義語展開指定の形式

同義語展開指定 ::=

SYNONYM 左括弧 同義語辞書名, 単純文字列指定 [, 展開モード指定] 右括弧

同義語展開指定時の注意事項

検索文字列にワイルドカードや除外文字検索用の特殊文字指定した場合は、同義語展開されません。

指定例

文字列「COMPUTER」の同義語、アルファベット大小文字および全角半角文字展開を含む（同義語辞書は USR01）。

```
SYNONYM( USR01, "COMPUTER", "AE" )
```

4.4.6 近傍条件指定

近傍条件指定では、二つの検索ターム（語句リスト）とそれらの距離条件、出現順序関係を指定します。それぞれの語句リスト中の検索タームのどれかがヒットした位置で、距離条件に当てはまるように近接している語句リストがヒットする文書を検索します。

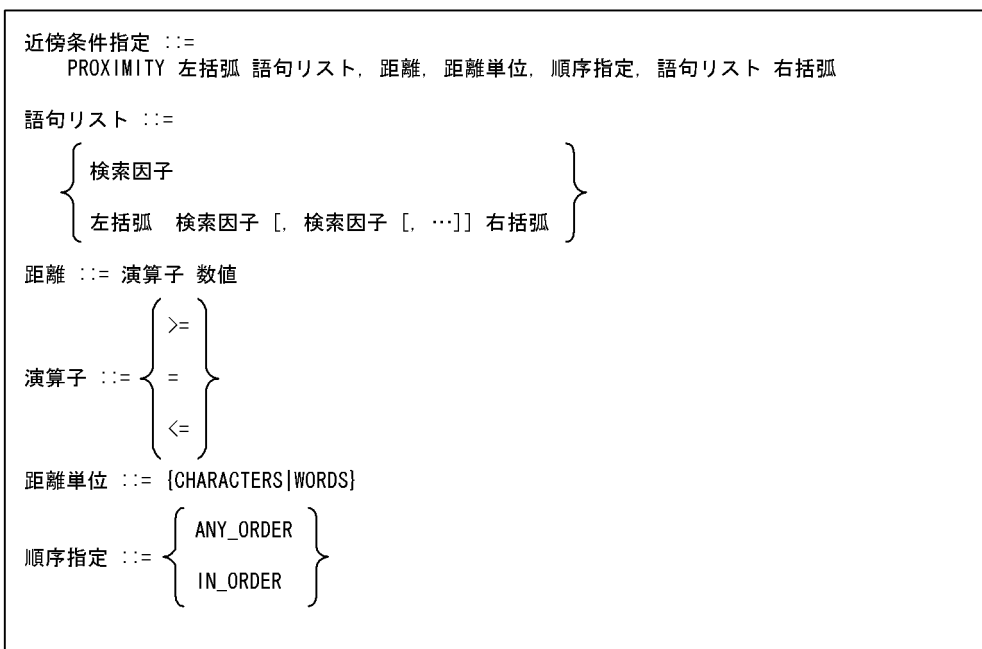
距離条件として、距離には演算子と数値、距離単位には文字を表す 'CHARACTERS' を使用します。これによって、二つの検索ターム間の文字数（距離）を検索条件にできます。

演算子は、指定距離に等しい・以上・以下の3種類があり、等しい場合は = を、以上の場合は >= を、以下の場合は <= を指定します。

順序指定は、第一の語句リストと第二の語句リストのどちらが先に出現してもよいとする 'ANY_ORDER' と、第一の語句リストの次に第二の語句リストが出現する順序を指定する 'IN_ORDER' を選択できます。

近傍条件指定の形式を次の図に示します。

図 4-22 近傍条件指定の形式



指定例

文字列「COMPUTER」の同義語，アルファベット大小文字および全角半角文字展開と，文字列「ネットワーク」の同義語，アルファベット大小文字，かな異表記，全角半角文字展開が 20 文字以下に近接して出現する（同義語辞書は USR01）。

```

PROXIMITY(SYNONYM( USR01, "COMPUTER", "AE"), <=20,
CHARACTERS, ANY_ORDER, SYNONYM( USR01, "ネットワーク", "AKE"))

```

- 英文検索用インデックスの場合

英文検索用インデックスは距離条件として，距離には演算子と数値，距離単位には単語を表す 'WORDS' を使用します。

英文検索用インデックスに対して距離単位に 'CHARACTERS' を指定した場合はエラーとなります。

近傍演算の距離は 0 ~ 2,147,483,646 の範囲で指定します。距離の単位は単語です。インデクスデリミタで区切られた単語を 1 単語とみなします。例えば，英文書に "I have a pen." とある場合，インデクスデリミタが半角スペースのため，'I' と 'pen' の距離は 2 となります。また，不要語が含まれる場合，不要語も 1 単語として近傍演算の距離に数えます。

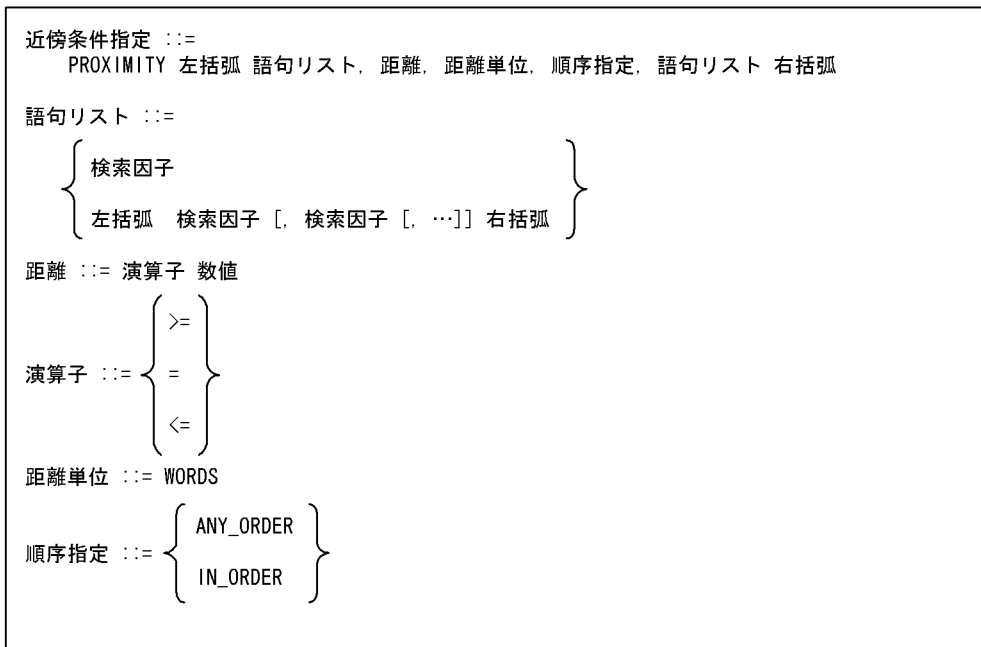
演算子は，指定距離に等しい・以上・以下の 3 種類があり，等しい場合は = を，以上の場合は >= を，以下の場合は <= を指定します。

順序指定は，第一の語句リストと第二の語句リストのどちらが先に出現してもよいとする 'ANY_ORDER' と，第一の語句リストの次に第二の語句リストが出現する順序を指定する 'IN_ORDER' を選択できます。

4. 抽象データ型関数を使用した文書または文字列データの操作

近傍条件指定の形式を次の図に示します。

図 4-23 英文インデクスの近傍条件指定の形式



指定例

文字列「COMPUTER」と「SYSTEM」が10単語以下に近接して出現する。

PROXIMITY("COMPUTER", <= 10, WORDS, ANY_ORDER, "SYSTEM")

4.4.7 特殊文字の指定

全文検索条件の条件式中で特殊な意味を持つ記号（特殊記号）は、単独では文字として指定できません。文字として指定する場合は、各特殊記号の前に"¥"（エスケープ文字）を付けてください。

全文検索条件の単純文字列指定以外と、単純文字列指定内では条件式中の特殊記号が異なります。

"¥"（エスケープ文字）を付ける必要のある特殊記号については、「4.3.12 特殊文字の指定」の表 4-8 および表 4-9 を参照してください。

4.4.8 全文検索条件式の指定例

全文検索条件式の指定例を次に示します。なお、例では表「reports」のFREEWORD型の列「fword」を検索対象としています。

文字列「COMPUTER」を含む文書の件数を調べる。

4. 抽象データ型関数を使用した文書または文字列データの操作

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{"COMPUTER"}' ) IS TRUE
```

文字列「COMPUTER」の同義語を含む文書のタイトル一覧表を作成する。

```
SELECT title FROM reports
WHERE contains(fword, '{SYNONYM(USR01, "COMPUTER")}' ) IS TRUE
```

文字列「COMPUTER」の同義語と「NETWORK」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{SYNONYM(USR01, "COMPUTER")
AND "NETWORK"}' ) IS TRUE
```

「COMPUTER」の同義語と「NETWORK」の同義語を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{SYNONYM(USR01,
"COMPUTER") AND SYNONYM(USR01,
"NETWORK")}' ) IS TRUE
```

「COMPUTER」の同義語と「NETWORK」の同義語が10文字以内に近接して出現する文書のタイトル一覧表を作成する。

```
SELECT title FROM reports
WHERE contains(fword, '{PROXIMITY(SYNONYM
(USR01, "COMPUTER"), <=10, CHARACTERS,
ANY_ORDER, SYNONYM(USR01,
"NETWORK"))}' ) IS TRUE
```

「COMPUTER」の同義語（同義語辞書USR01）と、「NETWORK」の同義語（同義語辞書POLITICS）が20文字以内に近接して出現する文書のタイトル、著者一覧表を作成する。

```
SELECT title, author FROM reports
WHERE contains(fword, '{PROXIMITY(SYNONYM
(USR01, "COMPUTER"), <= 20, CHARACTERS,
ANY_ORDER, SYNONYM( POLITICS,
"NETWORK"))}' ) IS TRUE
```

「COMPUTER」を含み、かつ「NETWORK」を含む文書の件数を調べる。

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{"COMPUTER"} AND {"NETWORK"}' ) IS TRUE
```

文字列「COMPUTER」と「NETWORK」のアルファベット大小文字展開と全角半角展開を含む文書の件数を調べる

```
SELECT COUNT(*) FROM reports
WHERE contains(fword, '{SOUNDEX_EXP("COMPUTER", "AE") AND
SOUNDEX_EXP("NETWORK", "AE")}' ) IS TRUE
```

4.5 概念検索条件式の文法

この節では、概念検索条件式の文法について説明します。

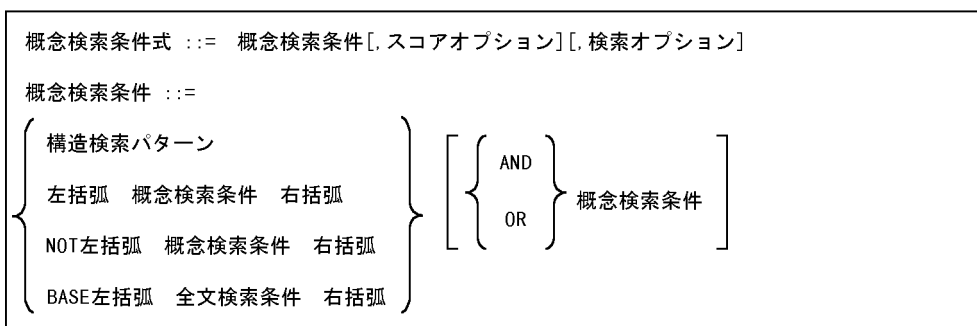
なお、concept_with_terms 関数を使用して特徴タームを抽出する際、概念検索条件式には、単純文字列指定だけを使用できます。

4.5.1 構造名を指定した検索条件の AND/OR 論理条件

概念検索条件には、構造検索パターンの AND/OR 論理条件を指定できます。また、概念検索条件に対して全文検索条件で絞り込み検索を指定することもできます。

概念検索条件式の形式を次の図に示します。

図 4-24 概念検索条件式の形式



注意事項

AND/OR 条件，または NOT 条件を指定する場合，各演算子について次の優先順位に従って演算します（左側の方が優先順位が高い）。

括弧 > NOT > AND > OR

概念検索条件に AND 条件と OR 条件を同じレベルで記述し，かつ OR 条件を優先したい場合は，OR 条件の範囲を括弧で囲んでください。

指定例

構造名「文章」の下の「概要」に，種文章「XXX」から抽出された検索用特徴タームを含む。

文章[概要{"XXX"}]

構造名「文章」の下の「概要」に，種文章「XXX」から抽出された検索用特徴タームを含む文書以外。

NOT(文章[概要{"XXX"}])

構造名「文章」の下の「概要」に，種文章「XXX」から抽出された検索用特徴タームの同義語を含む（同義語辞書は「USR01」）。

文章 [概要 {SYNONYM (USR01, "XXX")}]]

4.5.2 構造名の指定

構造検索パターンでは、検索対象とする文書要素をその構造名で指定します。また、構造指定のあとに、その文書要素内での検索条件を指定します。

構造名指定は、検索対象の構造名を文字列で指定します。

波括弧で囲まれたテキスト検索条件は、指定構造内での概念検索条件を指定する条件式です。

角括弧で構造検索パターンを囲むことで、上位の構造名指定と、指定した上位構造指定以下の構造名をさらに指定した複雑な構造指定ができます。これによって、例えば、要旨の下位構造の目的の部分で、テキスト検索条件を満たす文書を探せます。

構造検索パターンの形式を次の図に示します。

図 4-25 構造検索パターンの形式

構造検索パターン ::=

構造名指定 [… [構造名指定 [重み] 左波括弧 テキスト検索条件 右波括弧] …]

注意事項

- 構造名指定では使用できない文字があります。構造名指定で使用できない文字については、「4.3.12 特殊文字の指定」を参照してください。
- 構造名指定は 1,024 バイト以内で指定してください。
- 構造名指定は、最上位構造名から順に指定してください。最上位構造名だけを指定した場合、または構造名指定を省略した場合には、文書に含まれるすべての構造を対象に検索します。
- 次のように指定した構造名が未登録の場合はヒットしません。
 - ・ 指定した構造を持った文書が登録されていない場合
 - ・ 指定した構造名が間違っている場合
- 構造名に半角スペース (0x20) は指定できません。

指定例

構造名「文章」の下の「概要」の下の「目的」に、種文章「XXX」から抽出された検索用特徴タームを含む。

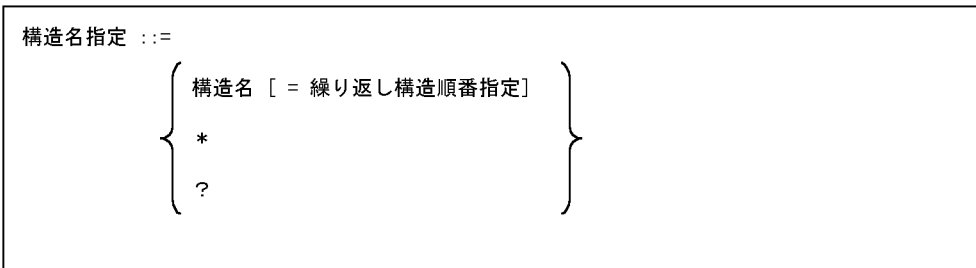
文章 [概要 [目的 {"XXX"}]]]

4.5.3 構造名のワイルドカード指定

構造名は、次の図に示す形式でワイルドカードが指定できます。

4. 抽象データ型関数を使用した文書または文字列データの操作

図 4-26 構造名指定の形式



* は、0 個以上の任意の構造階層を示します。

? は、1 個の任意の構造階層を示します。

また、構造名指定時に、繰り返して出現する構造名の番号を指定できます。「構造名 = 」のあとに、繰り返し構造の順番を指定します。指定方法は次のとおりです。

数 1

数 1 に示された順番を検索対象とします (順番の単独指定) 。

数 1 -

数 1 以降の順番を検索対象とします (順番の開始指定) 。

- 数 1

数 1 までの順番を検索対象とします (順番の最終指定) 。

数 1- 数 2

数 1 ~ 数 2 の順番を検索対象とします (順番の範囲指定) 。

数 1, 数 2, ……

指定された数 (数 1, 数 2) の順番を検索対象とします (任意の順番指定) 。

注意事項

- "*" (アスタリスク), "?" (疑問符), "=" (イコール), "-" (ハイフン), ",", " (コマ), および繰り返し構造順番指定の数字は、ASCII コードで指定します。
- 順番指定の範囲は 1 ~ 2,147,483,647 で指定します。
順番指定での範囲指定 "-" の値は、数 1- 数 2 の場合、数 1 < 数 2 の関係で指定します。数 1 が数 2 より大きい値の場合はエラーとなります。
- 順番指定での任意の指定 ", " の値に、同じ値を指定した場合、一つの設定とします。
例えば、「数 1, 数 2, 数 1」は「数 1, 数 2」となります。
- 順番指定での範囲指定 "- " と任意の指定 ", " は混在できません。エラーとなります。
例えば、「数 1- 数 2, 数 3」はエラーとなります。
- "*" は最下位に指定できません。

- "*" 直後の下位構造に "*", "?" は指定できません。
例えば、「*["?{" 検索ターム "}]」のような指定です。

指定例

構造名「文章」の下の「概要」の下の2番目の「コラム」に、種文章「XXX」から抽出された特徴タームを含む。

```
文章[概要[コラム=2[{"XXX"}]]
```

構造名「文章」の下の「概要」以下の「コラム」に、種文章「XXX」から抽出された特徴タームを含む。

```
文章[概要[*[コラム{"XXX"}]]]
```

構造名「文章」の下の「概要」の一つ下の階層に、種文章「XXX」から抽出された特徴タームを含む。

```
文章[概要[?["XXX"]]]]
```

4.5.4 テキスト検索条件の指定

テキスト検索条件は、検索因子と呼ぶ条件単位によって構成されています。検索因子を使用して、同一の文書構造内の複雑な検索条件を指定できます。

テキスト検索条件の形式を次の図に示します。

図 4-27 テキスト検索条件の形式

```
テキスト検索条件 ::= 検索因子
```

検索因子の形式を次の図に示します。

図 4-28 検索因子の形式

```
検索因子 ::=
```

```
{
  単純文字列指定
  異表記展開指定
  同義語展開指定
}
```

指定例

種文章「XXX」から抽出された検索用特徴タームの同義語を含む（同義語辞書はUSR01）

```
SYNONYM(USR01, "XXX")
```

4.5.5 単純文字列指定

単純文字列指定では、検索したい内容や概念を含む文字列（種文章）をダブルクォーテーションで囲んで指定します。

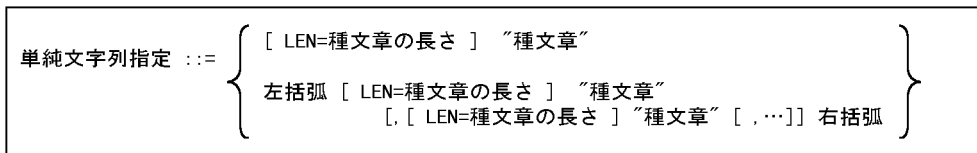
概念の対象となる文字の種類は、漢字、かな、およびアルファベットです。また、種文章に改行コードが含まれている場合は、改行コードを無視して、改行コードの前後は詰めて扱われます。

複数の種文章を指定する場合は、種文章をコンマで区切って、全体を括弧で囲んで指定します。この場合、それぞれの種文章の概念を包含した形で検索します。

種文章の長さは、ダブルクォーテーションで囲んだ文字列分（改行コードなども含む）をバイト単位で指定します。種文章の長さ指定を省略した場合、種文章に特殊記号が含まれているとエラーとなります。種文章の長さ指定を省略、かつ特殊記号を単なる文字として使用したい場合は、特殊記号の前に"¥"（エスケープ文字）を付けてください。単純文字列指定をする場合に、文字として指定できない特殊記号については、「4.3.12 特殊文字の指定」を参照してください。

単純文字列指定の形式を次の図に示します。

図 4-29 単純文字列指定の形式



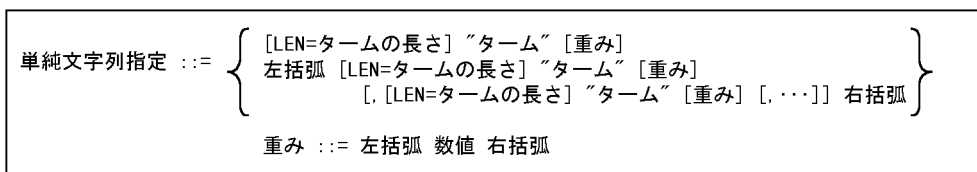
指定例

種文章「XXX」(長さ 100 バイト) から抽出された検索用特徴タームを含む。

```
LEN=100 "XXX"
```

検索オプションに term を指定することで、種文章ではなく、特徴タームを指定した検索ができます。検索オプションに term を指定した場合の単純文字列指定の形式を次の図に示します。

図 4-30 単純文字列指定の形式（検索オプションに term を指定した場合）



4.5.6 異表記展開指定

異表記展開指定では、検索条件となる種文章から抽出した検索用特徴タームと一致する文字列を検索するだけでなく、同音異表記の文字列も対象として検索します。例えば、検索用特徴タームが「ボウリング」であった場合、「ボウリング」だけでなく、「ボーリング」や「ポウリング」なども検索対象となります。このように、異表記展開を指定すると、より広い範囲での検索が可能となります。

異表記展開指定の形式を次の図に示します。

図 4-31 異表記展開指定の形式

異表記展開指定 ::=

SOUNDEX_EXP 左括弧 単純文字列指定 [. 展開モード指定] 右括弧

展開モードは、展開する異表記に応じて指定します。異表記オプションの種類と指定する展開モードの対応を次の表に示します。

表 4-11 異表記オプション

| 異表記オプション | 展開モード | 説明 |
|--------------|-------|--|
| アルファベット異表記 | A | 全角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |
| かたかな異表記 | K | 全角かたかな文字列に対して、かたかな異表記の展開をします。 |
| かたかな長音ハイフン展開 | L | かたかな文字列に対して、長音・ハイフン展開をします。このとき、展開されるのは長音・ハイフンだけであり、前後のかたかな文字列は展開の対象なりません。 |
| かたかなベーシック展開 | B | かたかなルールに基づいた展開だけを実行します。 |
| かたかな拗音展開 | C | かたかな文字列に対して、拗音展開をします。 |
| 全角半角異表記 | E | 指定した文字列中で全角から半角にできるもの、または半角から全角にできるものに対して、次の 3 パターンに展開します。 1. すべてが全角文字列 2. すべてが半角文字列 3. 指定した文字列 |
| 半角アルファベット展開 | H | 半角英字の文字列を、次の 4 パターンに展開します。 1. すべてが大文字 2. すべてが小文字 3. 先頭だけ大文字 4. 指定した文字列 |
| ひらがな拗音展開 | Y | ひらがなの拗音の展開をします。 |

4. 抽象データ型関数を使用した文書または文字列データの操作

| 異表記オプション | 展開モード | 説明 |
|-----------|-------|----------------------|
| 英単語派生表記展開 | S | 半角英文字の文字列を派生語に展開します。 |
| 英単語語尾表記展開 | I | 英単語を語尾の変換系に展開します。 |

注

このオプションを使用するには、Text Search Enhancer for English が必要です。異表記オプションの種類と指定する展開モードの対応の詳細については、表 4-7 を参照してください。

展開モードは単独で指定するだけでなく、組み合わせで指定することもできます。例えば、全角文字のアルファベット大小文字展開、および全角半角文字展開をする場合は、"AE" と指定します。展開モードの指定がない場合は、"AKE" が指定されていると仮定されます。なお、"BCL" を指定した場合は、"K" を指定したときと同じ展開となります。

展開モードは、順不同で指定できます。必要な組み合わせを指定してください。

なお、異表記展開の例については、「4.3.6 異表記展開指定」を参照してください。

指定例

種文章「XXX」から抽出された検索用特徴タームの、アルファベット大小文字および全角半角展開を含む。

```
SOUNDEX_EXP( "XXX", "AE" )
```

4.5.7 同義語展開指定

同義語展開指定では、同義語展開に使用する辞書名と種文章を指定します。検索条件となる種文章から抽出した検索用特徴タームを同義語展開することで、概念検索を実行します。

また、異表記展開モードの指定もできます。異表記展開モードを指定した場合は、同義語展開したすべての語に関して、指定のモードで異表記展開します。異表記展開モードの指定がない場合は、同義語展開の結果で検索します。

同義語展開指定の形式を次の図に示します。

図 4-32 同義語展開指定の形式

同義語展開指定 ::=

SYNONYM 左括弧 同義語辞書名. 単純文字列指定 [、展開モード指定] 右括弧

指定例

種文章「XXX」から抽出された検索用特徴タームの同義語、アルファベット大小文字および全角半角展開を含む（同義語辞書はUSR01）。

```
SYNONYM( USR01, "XXX", "AE" )
```

4.5.8 重み

重みは、検索対象である構造に対して付けます。複数構造に対して検索を実行する場合に、重要度を高くしたい構造に対して重みを付けます。これによって、重みを付けた構造にはより重要度の高いスコアが付きまます。重みを指定できる範囲は、1 ~ 100 です。

重み指定の形式を次の図に示します。

図 4-33 重み指定の形式

| |
|-------------------|
| 重み ::= 左括弧 数値 右括弧 |
|-------------------|

指定例

構造名「文章」の下の「本文」に、種文章「XXX」から抽出された検索用特徴タームを含み、かつ構造名「文章」の下の「概要」に、種文章「YYY」から抽出された検索用特徴タームを含む。

スコア値を算出する際は、種文章「XXX」(の検索用特徴ターム)に重み「10」を付ける。

文章[本文(10){"XXX"}] AND文章[概要{"YYY"}]

4.5.9 スコアオプション

スコアオプションでは、スコア付けした文書に対して次の指定ができます。

種文章に対して算出したスコアを基準(100点)として、検索結果の文書のスコアを相対的な値に正規化する指定

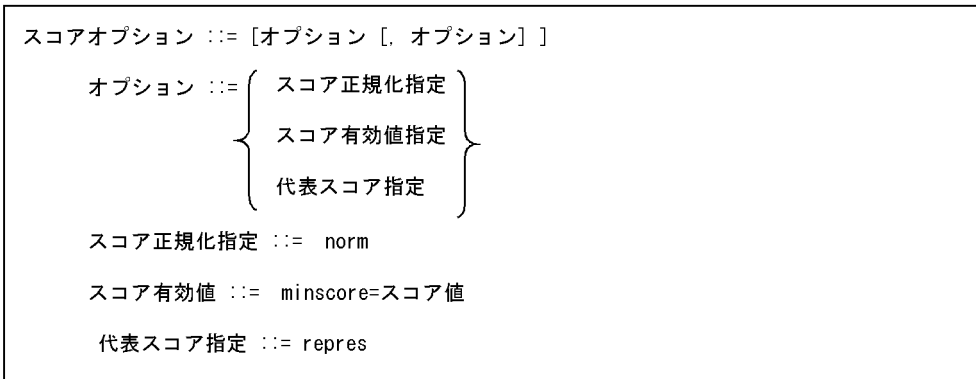
検索結果の件数を絞るためにスコアの下限値の指定

繰り返し構造に対するスコアの算出法の指定

なお、種文章に対して算出したスコアを基準とした場合(スコア正規化指定の場合)に、検索結果の中で種文章よりも適合度が高い文書のスコアは、種文章のスコアよりも高くなります。

スコアオプション指定の形式を次の図に示します。

図 4-34 スコアオプション指定の形式



適合度（スコア）の評価（算出）は、種文章から抽出された各検索用特徴タームに対して、次のような計算上の重みを付けて行います。

種文章とデータベースに登録されている文書の情報に基づいて算出された重み（Text Search Plug-in の MASTER インデクスに登録されている文書の情報が使われます）

各検索用特徴タームがヒットした文書中出现する割合に基づいて算出された重み

スコアの取得は、次の二つを用途によって使い分けます。

スコア正規化指定

種文章に対して算出したスコアを基準（100点）とし、検索結果の文書のスコアを正規化します。検索結果には、100点を超える文書もあります。

また、正規化時には小数点以下のスコア値を切り捨てるため、結果としてスコア値が0点となる場合があります。

スコア有効値指定

指定されたスコア以上の文書だけを検索結果とします。1 ~ 2,147,483,647 の値を指定できます。

ただし、NOT 条件を含む検索実行時にはスコア値はナル値となるため、スコア有効値を指定しても無視されます。

指定例

構造名「文章」の下の「概要」に、種文章「XXX」から抽出された検索用特徴タームを含む。スコアは正規化によって算出し、スコア値が60以上のものを対象とする。

文章[概要{"XXX"}],norm,minscore=60

また、登録文書に繰り返し構造が多い場合は、次のオプションの指定を検討してください。

代表スコアオプション指定

構造化文書での概念検索のスコア値は、ヒットした構造すべてのスコア値の合計と

なります。このため、繰り返し構造を指定した検索の場合、繰り返し構造数の多い文書のスコア値が極端に高くなる場合があります。このオプションを指定すると、ヒットした構造の中で最もスコア値が高い構造のスコア値を代表として、その文書のスコア値にできます。

なお、このオプションは、検索対象が繰り返し構造でない場合でも有効です。

指定例

繰り返し構造「章」を対象に代表スコア指定で概念検索する
書誌[章{"種文章"}], repres

4.5.10 検索オプション

検索オプションに term を指定することで、単純文字列指定に検索用特徴タームを指定して検索できます。また、複数の検索用特徴タームを指定する場合は、タームごとに重みを指定できます。重みを指定できる範囲は、1 ~ 100 です。指定を省略した場合は、重みを 1 として検索します。

なお、指定できる検索用特徴タームの数は、n-gram インデクスプラグインの環境設定項目「検索用特徴ターム数」に制限されません。検索オプションに term を指定した場合は、スコアオプションのスコア正規化指定 norm は無効になります。

検索オプション指定の形式を次の図に示します。

図 4-35 検索オプション指定の形式

検索オプション ::= term

4.5.11 絞り込み指定

概念検索に「AND BASE (全文検索条件)」を指定することで、概念検索の実行結果に対して絞り込み検索ができます。このとき、BASE 中の全文検索条件のスコア値は算出されません。また、概念検索条件中または全文検索条件中に NOT 条件を指定した場合は、スコア値はすべてナル値となります。

指定例

- 種文章「XXX」の概念検索実行結果に対して、検索ターム「AAA」で絞り込む。

```
{"XXX"} AND BASE ({"AAA"})
```

- 種文章「XXX」の概念検索実行結果に対して、OR 条件を指定した全文検索（検索ターム「AAA」および「TE*」）で絞り込む。

```
{"XXX"} AND BASE ({"AAA" OR "TE*"})
```

- 種文章「XXX」の概念検索実行結果に対して、近傍条件指定（文字列「日立」と文字列「製作所」が 10 文字以内で出現する）の全文検索で絞り込む。

4. 抽象データ型関数を使用した文書または文字列データの操作

```
{ "XXX" } AND BASE ( { PROXIMITY ("日立", <=10, CHARACTERS, ANY_ORDER, "製作所") } )
```

4.5.12 概念検索条件式の指定例

概念検索条件式の指定例を次に示します。例では表「reports」の SGMLTEXT 型の列「doc」を検索対象としています。また、「概要」および「目的」は、最上位構造「文章」の低位構造です。

なお、この項で説明する例では、概念検索条件式の指定例をわかりやすく示すため、SQL 文中に条件式をそのまま記述しています。

「概要」部分を対象に、種文章と似た概念を持つ文書を検索する。その際、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[概要{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}]') IS TRUE ORDER BY 1 DESC
```

「概要」部分を対象に、複数の種文章と似た概念を持つ文書を検索する。その際、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[概要{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}]') IS TRUE ORDER BY 1 DESC
```

「概要」部分を対象に、種文章から抽出した検索用特徴タームについて同義語展開して検索し、ヒットした文書のスコア、タイトル一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[概要{SYNONYM(USR01, "概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。")}]) IS TRUE ORDER BY 1 DESC
```

「目的」部分と「概要」部分を対象にそれぞれ種文章を指定して、種文章に似た概念を持つ文書を検索し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[目的{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}] AND文章[概要{"構造名指定による検索範囲の指定ができる。"}]') IS TRUE
ORDER BY 1 DESC
```

「文章」全体を対象に、種文章に似た概念を持つ文書を検索し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}')
```

```
) IS TRUE ORDER BY 1 DESC
```

「文章」の下階層にある 2 番目の「コラム」部分を対象に、種文章に似た概念を持つ文書を検索し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[コラム=2{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}]) IS TRUE ORDER BY 1 DESC
```

「概要」部分を対象に、種文章に似た概念を持つ文書を検索し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。なお、ヒットした文書には正規化したスコアを付け、スコア値が 60 以上の文書だけを検索結果とする。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[概要{"概念検索とは、検索条件として入力された文章または文字列と概念(内容)が似通った文書を検索する機能である。"}],norm, minscore=60') IS TRUE ORDER BY 1 DESC
```

「概要」部分を対象に、種文章に似た概念を持つ文書を検索し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。なお、種文章には特殊文字(", [,], =)が含まれるため、種文章の長さ(バイト数)を指定する。

```
SELECT score_concept(doc), title FROM reports
WHERE concept_with_score(doc, '文章[概要{LEN=99 "概念構造検索とは、検索条件として入力された"種文章"と概念[=内容]が似通った文書を検索する機能である。"}]) IS TRUE ORDER BY 1 DESC
```

種文章から検索用特徴タームを取得する。

```
SELECT terms_concept(doc) FROM reports WHERE
concept_with_terms(doc, '{"概念検索とは検索条件として入力された文章または文字列と概念(内容)が似通った文章を検索する機能である。"}') IS TRUE
```

種文章 A から抽出した検索用特徴タームについて、"似"と"通"と"文書"というタームを削除、"類似"というタームを重み 50 で追加して、ターム入力によって「概要」部分を対象に概念検索を実行する。その際、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
[種文章 A から抽出した検索用特徴タームと重みの例]
"検索"(100),"条件"(82),"入力"(75),"文章"(55),"概念"(54),"似"(43),"通"(40),"文書"(39)
SELECT score_concept(doc), title FROM reports
WHERE concept_with_score(doc, '文章[概要{"検索"(100),"条件"(82),"入力"(75),"文章"(55),"概念"(54),"類似"(50)}],term') IS TRUE ORDER BY 1 DESC
```

「概要」部分を対象に、複数の種文章と似た概念を持つ文書の中から、"論理条件"または"慣用句"を含む文書を検索する。その際、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc), title FROM reports
WHERE concept_with_score(doc, '文章[概要{"概念検索とは、検索条件とし
```

4. 抽象データ型関数を使用した文書または文字列データの操作

て入力された文章または文字列と概念（内容）が似通った文書を検索する機能である。", "構造名指定による検索範囲の指定ができる。また、検索条件同士のAND/OR論理演算ができる。")}] AND BASE(文章[概要{"論理条件" OR "慣用句"}])') IS TRUE ORDER BY 1 DESC

「概要」部分を対象に、種文章と似た概念を持つ文書を検索する。その際、繰り返し構造「概要」を対象に代表スコア指定し、スコア順にソートし、スコア、タイトルを格納する一覧表を作成する。

```
SELECT score_concept(doc),title FROM reports
WHERE concept_with_score(doc, '文章[概要{"概念検索とは、検索条件として入力された文章または文字列と概念（内容）が似通った文書を検索する機能である。"}],repres' ) IS TRUE ORDER BY 1 DESC
```


5

推奨する運用方法

この章では、Text Search Plug-in の推奨する運用方法について説明します。

5.1 推奨するデータ登録の運用方法

5.2 性能向上の運用

5.3 ログおよびバックアップの取得方式

5.4 概念検索用統計情報の収集，統合および反映

5.1 推奨するデータ登録の運用方法

5.1.1 登録運用の種類

Text Search Plug-in で提供している機能を利用して、次のような運用方法でデータを登録できます。

UAP の更新による運用

INSERT 文または UPDATE 文を使用し、UAP でデータを登録する方法です。

バッチ登録による運用

HiRDB の「pload (データベース作成ユーティリティ)」を使用してデータを登録する方法です。

プラグインインデクスの遅延一括作成による運用

インデクスを作成しないで UAP でデータを登録します。UAP の終了後、一括してインデクスを作成する方法です。

Text Search Plug-in は、プラグインインデクスの遅延一括作成機能をサポートしています。プラグインインデクスの遅延一括作成については、マニュアル「HiRDB システム運用ガイド」を参照してください。

差分インデクスによる運用

UAP でデータを登録したあと、別に確保した小容量の更新用エリアにインデクスを作成する方法です。

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、「phnstartidx または phxstartidx (インデクス作成開始ユーティリティ)」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、ユー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。

統一インデクスによる運用

大小文字・全角半角文字または拗音を区別しないで検索したい場合に、区別しない文字を統一したインデクスを作成する方法です。

また、これらの登録運用に加えて、データの登録と検索を同時に実行する運用方法があ

ります。

登録・検索同時実行による運用

排他制御の単位を小さくすることで、SQL による登録と検索の同時実行を可能とする運用

登録運用の長所および短所を表 5-1 に、登録・検索同時実行による運用の長所および短所を表 5-2 に示します。

表 5-1 Text Search Plug-in の登録運用の長所および短所

| 方式 | 長所 | 短所 |
|------------------------|--|--|
| UAP の更新による運用 | INSERT 文または UPDATE 文による通常の更新運用ができます | 登録済みのデータの増大に伴い、登録時間および登録単位のログ量が増大します |
| バッチ登録による運用 | <ul style="list-style-type: none"> 一括してデータ更新ができます 1 件ずつデータ登録するよりもログ量が削減できます | なし |
| プラグインインデクスの遅延一括作成による運用 | <ul style="list-style-type: none"> UAP のスループットが通常の更新に比べて向上します ログ量はバッチ登録と同等です | <ul style="list-style-type: none"> 登録データの検索がインデクスを作成するまで実行できません インデクスは別途作成する必要があります。この間、処理中の SGMLTEXT 型および FREEWORD 型に対する検索およびデータ更新はできません UPDATE 文を実行した場合、元データは削除されるため、検索対象外となります |
| 差分インデクスによる運用 | <ul style="list-style-type: none"> UAP のスループットが通常の更新に比べて向上します。また、インデクスも作成するため検索対象にできます ログ量は通常の場合に比べて削減できます | <ul style="list-style-type: none"> MASTER インデクスへの反映処理を別途する必要があります。この間、処理中の SGMLTEXT 型に対する検索およびデータ更新はできません MASTER インデクスに反映する際に、ログが出力されます FREEWORD 型では使用できません |

5. 推奨する運用方法

| 方式 | 長所 | 短所 |
|-----------------------------------|--|--|
| 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作成をする運用 | <ul style="list-style-type: none"> インデクス作成時の処理時間が短くなります インデクスへの文書または文字列データの登録を時間を区切って実行できます ログ量はバッチ登録と同じです | <ul style="list-style-type: none"> インデクス未登録状態の文書または文字列データは、検索対象になりません インデクス作成開始ユティリティを実行し、インデクスへ文書または文字列データを登録する必要があります インデクス未登録状態の文書または文字列データは、インクリメンタルの対象外になります |
| 分割遅延登録による運用 | <ul style="list-style-type: none"> UAPのスループットが通常の更新と比べて向上します ログ容量はバッチ登録と同等です インデクス作成時に、時間と件数を指定できます 登録・検索同時実行と併用できます | <ul style="list-style-type: none"> 登録データの検索がインデクスを作成するまで実行できません UPDATE文を実行した場合、元データは削除されるため検索対象外となります |
| 統一インデクスによる運用 | <ul style="list-style-type: none"> 異表記展開と同様の検索が高速にできます。 同義語異表記展開用メモリを削減できます。 | <ul style="list-style-type: none"> 大小文字・全角半角文字または拗音を区別した検索ができなくなります。 |

注

1件ずつ登録する場合、ログ量は通常の更新と同じです。

表 5-2 登録・検索同時実行による運用の長所および短所

| 方式 | 長所 | 短所 |
|-------------------|--|--|
| 登録・検索同時実行による運用 | <ul style="list-style-type: none"> SQLによる単数の更新系トランザクションと、複数の検索系トランザクションを同時に実行できます インデクス未作成の場合、複数の更新系トランザクションを同時に実行できます | <ul style="list-style-type: none"> ほかのユーザが更新中（トランザクションが終了していない状態）の文書または文字列データを検索できます Text Search Plug-inのデータの整合性を確保する排他制御のため、オーバーヘッドが掛かります。また、排他制御が競合した場合、一時的に待ち時間が発生します 共用メモリが必要です ロールバックによる回復処理中のRDエリアに対して検索した場合、「回復中であるため、検索できない」というエラーになる場合があります |
| 登録・検索同時実行を使用しない運用 | <ul style="list-style-type: none"> Text Search Plug-inのデータの整合性確保のための排他制御によるオーバーヘッドが掛かりません 共用メモリは不要です | <ul style="list-style-type: none"> 更新系トランザクションと、ほかのトランザクションを同時に実行できません |

5.1.2 推奨する登録運用

Text Search Plug-in の推奨する登録運用には、次の四つの方法があります。

- UAP の更新およびバッチ登録による運用
- プラグインインデクスの遅延一括作成による運用
- 差分インデクスによる運用
- 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

また、これらに加えて、次の二つの登録運用があります。

- 登録・検索同時実行による運用
- 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

以降、各運用方法について説明します。

(1) UAP の更新およびバッチ登録による運用の場合

(a) 差分インデクスの定義

差分インデクスを定義しないで、インデクス定義をします。ただし、あとで必要になる場合は定義しておいてください。

(b) 運用方法

通常の HiRDB の運用に準じます。

(2) プラグインインデクスの遅延一括作成による運用の場合

(a) 差分インデクスの定義

バッチ登録および UAP の更新時に差分インデクスを使用しない場合、差分インデクスの定義は不要です。

(b) 運用方法

UAP でバッチ運用をする場合に推奨します。

削除要求 (UPDATE 文または DELETE 文) は即時反映しますが、追加要求は即時反映をしないためデータの整合性が確保できません。したがって、検索運用中の使用は推奨しません。

プラグインインデクスの遅延一括作成については、マニュアル「HiRDB システム運用ガイド」を参照してください。

(3) 差分インデクスによる運用の場合

(a) 差分インデクスの定義

差分インデクスを定義してください。

5. 推奨する運用方法

(b) 運用方法

UAP で通常の更新をする場合に使用を推奨します。

ただし、次の手順での運用を前提とします。

1. INSERT 文または UPDATE 文を使用してデータを登録する
2. 運用の合間に、差分インデクスの情報を MASTER インデクスへ反映する
この処理中の注意点については、「付録 C 資源の排他」を参照してください。

注意

- 最初のデータは、定義内容に関係なく、MASTER インデクスに登録します。
- 差分インデクスの情報を MASTER インデクスに反映するタイミングは、テキスト容量で約 10 メガバイトを目安にしてください。

(4) 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の場合

(a) インデクス作成方法指定の定義

インデクス作成方法指定に初期分割登録を定義してください。

(b) 運用方法

表に大量の文書または文字列データを登録したあと、限られた時間単位でインデクスへ文書を登録する場合に使用を推奨します。「phnstartidx または phxstartidx (インデクス作成開始ユティリティ)」によって、すべての文書または文字列データがインデクスに登録されるまでは、データの整合性を確保できません。したがって、すべての文書または文字列データをインデクスに登録する前に検索を実行する場合には、使用を推奨しません。

注意

- インデクス未登録状態の文書または文字列データがなくなれば、UAP によるデータ更新の運用、バッチ登録による運用、プラグインインデクスの遅延一括作成による運用、差分インデクスによる運用に移行できます。差分インデクスによる運用に移行するには、インデクス作成時に差分インデクスの定義が必要です。

(5) 登録・検索同時実行による運用の場合

(a) 登録・検索同時実行の定義

登録・検索同時実行を定義してください。

(b) 運用方法

SQL による更新系トランザクションが、検索系トランザクションと同時に実行される環境での使用を推奨します。

注意

- インデクス未作成の場合の更新系トランザクション，または HiRDB のプラグインインデクスの遅延一括作成を使用している場合の INSERT 文によって登録する更新系トランザクションは複数同時に実行できますが，それ以外の更新系トランザクションは複数同時に実行できません。
- HiRDB が提供しているユティリティによる登録と検索系トランザクションは同時に実行できません。また，Text Search Plug-in の一部のユティリティと検索系トランザクションは同時に実行できません。詳細は「3. Text Search Plug-in のユティリティ」を参照してください。
- n-gram インデクスプラグインを使用している場合は，n-gram インデクス LOB 用 RD エリアに対してグローバルバッファを指定してください。また，IXFREEWORD プラグインを使用している場合は，IXFREEWORD インデクス LOB 用 RD エリアに対してグローバルバッファを指定してください。グローバルバッファを指定しない場合に比べて，登録性能が向上します。
- 同時実行を設定している場合，Text Search Plug-in のデータの整合性確保のために，一時的に排他制御をします。このため，同時実行を設定していない場合に比べて，排他制御によるオーバーヘッドが掛かります。
- ほかのユーザが更新中（トランザクションが終了していない状態）の文書または文字列データを検索できます。

(6) 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の場合

(a) 分割遅延登録用インデクスの定義

分割遅延登録用インデクスを定義するには，次の方法があります。

- 定義系 SQL の CREATE INDEX によって定義します。
- 「phnmodidx または phxmodidx（インデクス情報変更ユティリティ）」によって，インデクス定義を分割遅延登録用インデクスに変更します。

(b) 運用方法

UAP でバッチ運用する場合に推奨します。

1. INSERT 文または UPDATE 文を使用してデータを登録します。
2. 運用の合間に，遅延されている情報を MASTER インデクスへ反映します。
インデクスの作成は，「phnstartidx または phxstartidx（インデクス作成開始ユティリティ）」を使用して実行します。
3. 検索処理とインデクス作成の同時実行する場合は，ユティリティ実行時に同時実行指定（-w nowait）を指定します。また，サーバ側の定義を登録・検索同時実行で定義します。
4. 分割遅延登録の開始 / 終了
即時更新 / 分割遅延登録の切り替えは，「phndelay または phxdelay（インデクス分割遅延登録運用変更ユティリティ）」を使用します。運用形態に合わせて選択してください。

5. 推奨する運用方法

注意

- 削除要求（UPDATE 文または DELETE 文）は即時反映します。しかし、追加要求は即時反映をしないため、データの整合性が確保できません。そのため、インデクス未作成文書に対して検索が必要である場合は推奨しません。
- 分割遅延登録用領域使用率が 100% になると、それ以上分割遅延登録を実行できません。100% にならないように、インデクスへ反映してください。インデクスへの反映を時間・件数で中断する場合は、使用率は下がりにません。使用率を下げるためには、すべての未反映文書または文字列をインデクスに反映してください。分割遅延登録用領域を使い切った場合の SQL コマンドについては、環境定義ファイルによって、エラーリターン / 即時更新を選択できます。
- 即時更新へ切り替える際に、インデクス未登録文書があるときはすべて登録します。時間・件数による中断はできないので注意してください。

(7) 統一インデクスによる運用の場合

(a) 統一インデクスの定義

大文字と小文字、または全角文字と半角文字の違いを区別しないで検索を実行したい場合には、大小文字・全角文字半角文字統一指定を定義してください。また、検索時に「や」と「ゃ」、「よ」と「ょ」などの拗音を区別しないで検索を実行したい場合には、拗音統一指定を定義してください。「バ」と「ハ」、「ば」と「ば」などの濁音や半濁音を区別しないで検索を実行したい場合には、濁音統一指定を定義してください。

(b) 運用方法

アルファベット大小文字展開や全角半角文字展開を必ず指定する運用の場合は、大小文字・全角文字半角文字統一指定での運用を推奨します。また、かたかな異表記展開やひらがな拗音展開を必ず指定する運用の場合は、拗音統一指定での運用を推奨します。

注意

- 大小文字・全角半角文字または拗音を区別した検索を実行したい場合は、インデクスを再作成する必要があります。

(8) 推奨する組み合わせ

登録運用を実行する場合の推奨するパラメタの組み合わせを次の表に示します。

表 5-3 登録運用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 | wait | off | wait |
| 2 | CREATEMODE=2 | nowait | off | nowait |
| 3 | CREATEMODE=3 | nowait_with_update | off | nowait_with_update |

1 に適した業務

n-gram インデクス作成または更新をバッチで実行し、かつ n-gram 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

n-gram インデクス作成または更新をバッチで実行し、かつ n-gram 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。

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、n-gram 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。

5.2 性能向上の運用

この節では、性能向上を目的とした機能、およびその運用方法について説明します。

5.2.1 グローバルバッファ管理の運用

Text Search Plug-in のデータや、インデクス格納用 RD エリアにグローバルバッファを指定する運用を行うと、データの入出力でシステムにかかる負荷を削減できます。

また、インデクス用グローバルバッファ管理方式は最適化指定ができます。最適化指定をすると、インデクスの管理情報を優先的にグローバルバッファに残すことで、最適化指定をしない場合に比べて、グローバルバッファを効率的に使用できます。グローバルバッファサイズを 100 メガバイト以上確保する場合は、最適化指定を行う運用を推奨します。

5.3 ログおよびバックアップの取得方式

この節では、ログ取得の方法およびバックアップを取得する方法について説明します。

5.3.1 登録運用のログ取得方式の設定について

登録運用のログ取得方式には、次の三つのモードがあります。ログ量は、あらかじめログ量の見積もりをして、不足しないように運用してください。

- ログ取得モード
- 更新前ログ取得モード
- ログレスモード

各モードの推奨する運用方法について次に説明します。

(1) ログ取得モード

UAP から登録する場合に推奨します。なお、ログ量を削減するため、差分インデクス機能を使用した運用を推奨します。

また、ログ取得モードを選択する場合は、表定義の列回復制約に ALL を指定してください。詳細については、マニュアル「HiRDB SQL リファレンス」を参照してください。

(2) 更新前ログ取得モード

HiRDB の「pload (データベース作成ユティリティ)」使用時、遅延更新機能および差分インデクスから MASTER インデクスへの反映をする場合に推奨します。

更新が終わったあとは、必ずバックアップを取得してください。バックアップは、HiRDB の「pdcopy (データベース複写ユティリティ)」で取得してください。

(3) ログレスモード

大量データに対するインデクス一括作成の場合に推奨します。ログを取得しない運用では、障害が発生した場合、データが回復されません。必ず、正常終了後にバックアップを取得してください。また、障害発生に備え、HiRDB のログレス運用の説明に従い、運用してください。

なお、検索・登録同時実行では、ログレスモードでは運用できません。

5.3.2 バックアップ取得方法

(1) データのバックアップ

データのバックアップは、HiRDB の「pdcopy (データベース複写ユティリティ)」で取得してください。pdcopy の使用方法は、マニュアル「HiRDB システム運用ガイド」および「HiRDB コマンドリファレンス」を参照してください。

(2) 運用環境のバックアップ

ここでは、Text Search Plug-in の運用環境のバックアップについて説明します。

Text Search Plug-in の運用環境のバックアップを行う場合は、HiRDB の各種設定に加えて Text Search Plug-in として以下の情報のバックアップが必要です。

(a) 同義語辞書

同義語辞書のバックアップは、OS により次のどちらかの方法で行ってください。

UNIX の場合

同義語辞書バックアップユーティリティを使用します。同義語辞書バックアップユーティリティの詳細については、SGMLTEXT 型の場合は「3.2.8 同義語辞書バックアップユーティリティ」を参照してください。また、FREEWORD 型の場合は「3.3.4 同義語辞書バックアップユーティリティ」を参照してください。

Windows の場合

次に示す同義語辞書データをコピーしてください。

- SGMLTEXT 型

```
%PDDIR%\plugin\_%phsgml%\dic
```

- FREEWORD 型

```
%PDDIR%\plugin\_%phcfwd%\dic
```

注

ディレクトリ名称が dic で始まるすべてのディレクトリ、およびディレクトリ内ファイルをコピーします。

(b) インデクス定義ファイル

次のディレクトリにインデクス定義ファイルが存在する場合は、コピーしてください。

UNIX の場合

- SGMLTEXT 型

```
$PDDIR/plugin/_phngram/idx_def
```

- FREEWORD 型の場合

```
$PDDIR/plugin/_phxfwr/idx_def
```

- 英文デリミタファイル
- 英文削除コードファイル
- 英文不要語ファイル

Windows の場合

- SGMLTEXT 型

```
%PDDIR%\plugin\_%phngram%\idx_def
```

- FREEWORD 型

```
%PDDIR%\plugin\_%phxfwr%\idx_def
```

- 英文デリミタファイル
- 英文削除コードファイル
- 英文不要語ファイル

(c) 環境定義ファイル

次のディレクトリ，およびディレクトリ内のファイルをコピーしてください。

UNIX の場合

- \$PDDIR/plugin/_phsgml/conf
- \$PDDIR/plugin/_phngram/conf
- \$PDDIR/plugin/_phcfwd/conf
- \$PDDIR/plugin/_phxfwr/conf

Windows の場合

- %PDDIR%\plugin_phsgml\conf
- %PDDIR%\plugin_phngram\conf
- %PDDIR%\plugin_phcfwd\conf
- %PDDIR%\plugin_phxfwr\conf

5.4 概念検索用統計情報の収集，統合および反映

横分割した表に文書を登録した場合，横分割した表単位に定義された概念検索用インデクスは，それぞれ異なる統計情報を保持します。このため，横分割した表から，検索用特徴タームを抽出すると，横分割した表ごとに異なる検索用特徴タームが抽出されることがあります。

横分割したそれぞれの表の検索用特徴タームを一意にするには，横分割した表ごとに異なる統計情報を収集，統合，そして反映します。統計情報を収集，統合し，反映するまでの手順を次に示します。なお，表を横分割しない運用の場合は，概念検索用統計情報は一意であるため，この手順は不要です。

1. 横分割した表のサーバ単位に，統計情報を出力するディレクトリを作成します。
2. 「`phnconget` (統計情報収集ユティリティ)」を実行して，1. で作成したディレクトリに統計情報を出力します。
このとき，`-o` オプションで指定する統計情報出力先定義ファイルには，1. で作成したディレクトリを指定してください。
3. 2. で出力した統計情報ファイルを「`phnconbuild` (統計情報統合ユティリティ)」を実行するマシンに集めます。
4. 「`phnconbuild` (統計情報統合ユティリティ)」を実行して，各統計情報を統合します。
5. 統合統計情報ファイルを，反映先の各サーバに配布します。
6. 「`phnconrstr` (統計情報反映ユティリティ)」を実行して，統合後の統計情報をデータベースに反映します。このとき，RD エリアに空き容量があるかどうかを確認してください。
この段階で，各 RD エリアの統計情報が一意になります。
7. 「`phnchgcon` (特徴ターム取得 RD エリア指定ユティリティ)」を実行して，特徴タームを返却する RD エリアを一つに限定します。
このとき，`-r` オプションで指定する RD エリアには，横分割したそれぞれの表に定義したインデクス用の RD エリアを指定してください。
このユティリティを実行しないで特徴ターム取得抽象データ型関数を使用すると，横分割した表の数だけ，検索用特徴タームが返却されます。

付録

付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字

付録 B ディレクトリおよびファイル構成

付録 C 資源の排他

付録 D 障害発生時の運用方法

付録 E 系切り替え

付録 F エラーメッセージ

付録 G 用語解説

付録 A 登録する文書（テキストデータ）および文字列データに使用できる文字

ここでは、Text Search Plug-in で扱う文書および文字列データに使用できる文字と、登録する文字に関する注意事項について説明します。

(1) 登録する文書（テキストデータ）および文字列データに使用できる文字

登録する文書（テキストデータ）および文字列データに使用できる文字を、表 A-1 および表 A-2 に示します。

表 A-1 登録する文書（テキストデータ）および文字列データに使用できる文字
(HiRDB の文字コード種別が SHIFT-JIS のとき)

| 文字種 | | 長さ (バイト) | 配置 | |
|-------|-----------------|----------|----------------------------|----------------------------|
| | | | 2 バイトコードの 1 バイト目 | 2 バイトコードの 2 バイト目 |
| 文字コード | 1 バイト 文字コード | 1 バイト | 0x20 ~ 0x7E 0xA1 ~ 0xDF | - |
| | 2 バイト 文字コード | 2 バイト | 0x81 ~ 0x9F 0xE0 ~ 0xFC | 0x40 ~ 0x7E 0x80 ~ 0xFC |
| 制御コード | JIS 制御 文字コード | 1 バイト | 0x09 | - |

表 A-2 登録する文書（テキストデータ）および文字列データに使用できる文字
(HiRDB の文字コード種別が EUC のとき)(UNIX の場合)

| 文字種 | | 長さ (バイト) | 配置 | |
|-------|-----------------|----------|---------------------|----------------------------|
| | | | 2 バイトコードの 1 バイト目 | 2 バイトコードの 2 バイト目 |
| 文字コード | 1 バイト 文字コード | 1 バイト | 0x20 ~ 0x7E | - |
| | 2 バイト 文字コード | 2 バイト | 0x8E 0xA1 ~ 0xFE | 0xA1 ~ 0xDF 0xA1 ~ 0xFE |
| 制御コード | JIS 制御 文字コード | 1 バイト | 0x09 | - |

(2) 英文で登録する文書（テキストデータ）および文字列データに使用できる文字

英文で登録する文書（テキストデータ）および文字列データに使用できる文字を、次の

表に示します。

表 A-3 登録する文書（テキストデータ）および文字列データに使用できる文字
（HiRDB の文字コード種別が SHIFT-JIS , EUC および utf-8 のとき）

| | 文字種 | 長さ（バイト） | 使用できる文字 |
|-------|-------------|---------|-------------|
| 文字コード | 1 バイト文字コード | 1 バイト | 0x20 ~ 0x7E |
| 制御コード | JIS 制御文字コード | 1 バイト | 0x09 |

なお，HiRDB の言語種別が utf-8 の場合は，Text Search Plug-in でも文字コード utf-8（MS-Unicode または JISX0221）を使用できます。ただし，インデクス定義に INDEX_UCS4 の指定はできません。

（3）utf-8 で使用できる文字コード

使用できる utf-8 の範囲を次の表に示します。

表 A-4 utf-8 コードで使用できる文字（表内の文字コード表現は 16 進数）

| コードポイント | 1 バイト | 2 バイト | 3 バイト | 4 バイト | 5 バイト | 6 バイト |
|--|-----------------------------|-------|-------|-------|-------|-------|
| U+00000009 , U+00000020 ~ U+0000007E | 09 , 20 ~ 7E | | | | | |
| U+00000080 ~ U+000007FF | C280 ~ DFBF | | | | | |
| U+00000800 ~ U+0000FFFF | E0A080 ~ EFBFBF | | | | | |
| U+00010000 ~ U+0010FFFF | F0908080 ~ F48FBFBF | | | | | |
| U+00110000 ~ U+001FFFFF | F4908080 ~ F7BFBFBF | | | | | |
| U+00200000 ~ U+03FFFFFF | F88808080 ~ FBBFBFBFBF | | | | | |
| U+04000000 ~ U+7FFFFFFF | FC8480808080 ~ FDBFBFBFBFBF | | | | | |

注

UCS4 の範囲の制御文字については，記述できない範囲があります。

（4）登録する文字に関する注意

（a）使用できない文字に関する注意

Text Search Plug-in では，可視状態でつながって見える文字でも，文字と文字の間に表 A-1 , A-2 および A-3 または A-4 に示す以外の文字が含まれていると，検索タームを指定

してもヒットしません。表 A-1, A-2 および A-3 または A-4 に示す以外の文字が登録文書中に含まれていても、登録は正常に処理されエラーにはならない場合があるため、表 A-1, A-2 および A-3 または A-4 に示す以外の文字が含まれているテキストデータの取り扱いにはご注意ください。

例えば、次のような場合は検索を実行してもヒットしません。

可視状態：日立

1 文字目に「日」、2 文字目に「立」と見えています。

実際のデータ：日<表 A-1 および A-2 に示す以外の文字>立

「日」と「立」の間に<表 A-1, A-2 および A-3 または A-4 に示す以外の文字>が含まれていると、1 文字目は「日」、2 文字目の「<表 A-1, A-2 および A-3 または A-4 に示す以外の文字>」は検索対象でないため無視され、3 文字目は「立」として登録されます。

検索：検索ターム「日立」

検索タームに「日立」を指定しても、実際のデータは「日」と「立」が連続していないためヒットしません。

また、<表 A-1, A-2 および A-3 または A-4 に示す以外の文字>は検索対象でないため、検索タームに<表 A-1, A-2 および A-3 または A-4 に示す以外の文字>を指定してもヒットしません。

「日<表 A-1, A-2 および A-3 または A-4 に示す以外の文字>立」をヒットさせたい場合は、検索タームとして「日?立」、「日??立」、「日*立」、「日」と「立」の近傍条件などを使用してください。

英文検索は、デリミタなどの定義によってはヒットする場合があります。

(b) 改行コードに関する注意

登録するテキストデータの改行コードには、LF (0x0a) を使用してください。登録するテキストデータに使用している改行コードが CR+LF (0x0d0a) の場合、改行コードを次のように処理します。

文書検索プラグインの文書登録種別が SGML の場合
0x0d0a と 0x0a を同じ改行コードとして処理します。

文書検索プラグインの文書登録種別が PLAIN の場合、および文字列検索プラグインの場合

0d 部分は表 A-1, A-2 および A-3 または A-4 に示した以外の文字として処理し、0a の部分は改行コードとして処理します。

付録 B ディレクトリおよびファイル構成

付録 B.1 ディレクトリおよびファイル構成 (UNIX の場合)

ここでは、Text Search Plug-in を HiRDB に登録したあとの、SGML プラグインと n-gram インデクスプラグインのディレクトリおよびファイル構成を表 B-1 に示します。また、FREEWORD プラグインと IXFREEWORD プラグインのディレクトリおよびファイル構成を次の表に示します。

表 B-1 SGML プラグインと n-gram インデクスプラグインのディレクトリおよびファイル構成 (UNIX の場合)

| ディレクトリ名 | ファイル名 | 説明 |
|-----------------------------|---------------|--------------------------------------|
| \$PDDIR/plugin/_phsgml | - | SGML プラグインを格納するディレクトリ |
| \$PDDIR/plugin/_phsgml/bin | - | SGML プラグインのユティリティを格納するディレクトリ |
| | phsregsetup | レジストリ初期化ユティリティ |
| | phssgmlreg | SGML 定義情報登録ユティリティ |
| | phssgmlrm | SGML 定義情報削除ユティリティ |
| | phssgmlval | SGML 定義情報参照ユティリティ |
| | phssyndicreg | 同義語辞書登録ユティリティ |
| | phssyndicconv | 同義語辞書移行ユティリティ |
| | phssyndicrm | 同義語辞書削除ユティリティ |
| | phssyndicbkup | 同義語辞書バックアップユティリティ |
| \$PDDIR/plugin/_phsgml/conf | - | SGML プラグインのコンフィグレーションファイルを格納するディレクトリ |

| ディレクトリ名 | ファイル名 | 説明 |
|---|-------------|------------------------------------|
| \$PDDIR/plugin/_phsgml/dic \$PDDIR/plugin/_phsgml/dic_c \$PDDIR/plugin/_phsgml/dic_w \$PDDIR/plugin/_phsgml/dic_y \$PDDIR/plugin/_phsgml/dic_cw \$PDDIR/plugin/_phsgml/dic_cy \$PDDIR/plugin/_phsgml/dic_wy \$PDDIR/plugin/_phsgml/dic_cwy \$PDDIR/plugin/_phsgml/dic_cd \$PDDIR/plugin/_phsgml/dic_cdr \$PDDIR/plugin/_phsgml/dic_cr \$PDDIR/plugin/_phsgml/dic_cwd \$PDDIR/plugin/_phsgml/dic_cwdr \$PDDIR/plugin/_phsgml/dic_cwr \$PDDIR/plugin/_phsgml/dic_cwyd \$PDDIR/plugin/_phsgml/dic_cwydr \$PDDIR/plugin/_phsgml/dic_cwyr \$PDDIR/plugin/_phsgml/dic_cyd \$PDDIR/plugin/_phsgml/dic_cydr \$PDDIR/plugin/_phsgml/dic_cyr \$PDDIR/plugin/_phsgml/dic_d \$PDDIR/plugin/_phsgml/dic_dr \$PDDIR/plugin/_phsgml/dic_r \$PDDIR/plugin/_phsgml/dic_wd \$PDDIR/plugin/_phsgml/dic_wdr \$PDDIR/plugin/_phsgml/dic_wr \$PDDIR/plugin/_phsgml/dic_wyd \$PDDIR/plugin/_phsgml/dic_wydr \$PDDIR/plugin/_phsgml/dic_wyr \$PDDIR/plugin/_phsgml/dic_yd \$PDDIR/plugin/_phsgml/dic_ydr \$PDDIR/plugin/_phsgml/dic_yr | - | 同義語辞書格納ディレクトリ |
| \$PDDIR/plugin/_phsgml/etc | - | SGML プラグイン登録用ファイル格納ディレ クトリ |
| | _phsgml.adt | SGML プラグイン定義ファイル |
| | _phsgml.pic | SGML プラグイン PIC ファイル |
| \$PDDIR/plugin/_phsgml/sample | - | SGML プラグインのサンプルファイルを格納す るディレクトリ |
| | bes01 | BES 単位指定コンフィグレーションファイル サンプルファイル |
| | pdplgsys | システム単位指定コンフィグレーションファイ ルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイ ルサンプルファイル |
| \$PDDIR/plugin/_phsgml/spool | - | SGML プラグインの障害情報格納ディレクトリ |
| \$PDDIR/plugin/_phsgml/tmp | - | 作業用ファイル格納ディレクトリ |
| \$PDDIR/plugin/_phngram | - | n-gram インデクスプラグインを格納するディ レクトリ |

| ディレクトリ名 | ファイル名 | 説明 |
|--------------------------------|--------------|---|
| \$PDDIR/plugin/_phngram/bin | - | n-gram インデクスプラグインのユーティリティを格納するディレクトリ |
| | phnidxls | インデクス情報の取得ユーティリティ |
| | phnincmt | インデクスのインクリメンタルユーティリティ |
| | phnmerge | インデクスのマージユーティリティ |
| | phnexpnd | インデクスの拡張ユーティリティ |
| | phnstartidx | インデクス作成開始ユーティリティ |
| | phncond | インデクス詰め替えユーティリティ |
| | phnmodidx | インデクス情報変更ユーティリティ |
| | phndelay | インデクス分割遅延登録運用変更ユーティリティ |
| | phnconget | 統計情報収集ユーティリティ |
| | phnconbuild | 統計情報統合ユーティリティ |
| | phnconstr | 統計情報反映ユーティリティ |
| | phnchgcon | 特徴ターム取得 RD エリア指定ユーティリティ |
| \$PDDIR/plugin/_phngram/conf | - | n-gram インデクスプラグインのコンフィグレーションファイルを格納するディレクトリ |
| \$PDDIR/plugin/_phngram/etc | - | n-gram インデクスプラグイン登録用ファイル格納ディレクトリ |
| | _phngram.idx | n-gram インデクスプラグイン定義ファイル |
| | _phngram.pic | n-gram インデクスプラグイン PIC ファイル |
| \$PDDIR/plugin/_phngram/sample | - | n-gram インデクスプラグインのサンプルファイルを格納するディレクトリ |
| | pdplgsys | システム単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイルサンプルファイル |
| | bes01 | BES 単位指定コンフィグレーションファイルサンプルファイル |
| \$PDDIR/plugin/_phngram/spool | - | n-gram インデクスプラグインの障害情報格納ディレクトリ |
| \$PDDIR/plugin/_phngram/tmp | - | 作業用ファイル格納ディレクトリ |

(凡例)

- : 該当しない。

表B-2 FREEWORD プラグインと IXFREEWORD プラグインのディレクトリおよびファイル構成 (UNIX の場合)

| ディレクトリ名 | ファイル名 | 説明 |
|---|---------------|--|
| \$PDDIR/plugin/_phcfwd | - | FREEWORD プラグインを格納するディレクトリ |
| \$PDDIR/plugin/_phcfwd/bin | - | FREEWORD プラグインのユティリティを格納するディレクトリ |
| | phcsyndicreg | 同義語辞書登録ユティリティ |
| | phcsyndicrm | 同義語辞書削除ユティリティ |
| | phcsyndiccon | 同義語辞書移行ユティリティ |
| | phcsyndicbkup | 同義語辞書バックアップユティリティ |
| \$PDDIR/plugin/_phcfwd/conf | - | FREEWORD プラグインのコンフィグレーションファイルを格納するディレクトリ |
| \$PDDIR/plugin/_phcfwd/dic \$PDDIR/plugin/_phcfwd/dic_c \$PDDIR/plugin/_phcfwd/dic_w \$PDDIR/plugin/_phcfwd/dic_y \$PDDIR/plugin/_phcfwd/dic_cw \$PDDIR/plugin/_phcfwd/dic_cy \$PDDIR/plugin/_phcfwd/dic_wy \$PDDIR/plugin/_phcfwd/dic_cwy \$PDDIR/plugin/_phcfwd/dic_cd \$PDDIR/plugin/_phcfwd/dic_cdr \$PDDIR/plugin/_phcfwd/dic_cr \$PDDIR/plugin/_phcfwd/dic_cwd \$PDDIR/plugin/_phcfwd/dic_cwdr \$PDDIR/plugin/_phcfwd/dic_cwr \$PDDIR/plugin/_phcfwd/dic_cwyd \$PDDIR/plugin/_phcfwd/dic_cwydr \$PDDIR/plugin/_phcfwd/dic_cwyr \$PDDIR/plugin/_phcfwd/dic_cyd \$PDDIR/plugin/_phcfwd/dic_cydr \$PDDIR/plugin/_phcfwd/dic_cyr \$PDDIR/plugin/_phcfwd/dic_d \$PDDIR/plugin/_phcfwd/dic_dr \$PDDIR/plugin/_phcfwd/dic_r \$PDDIR/plugin/_phcfwd/dic_wd \$PDDIR/plugin/_phcfwd/dic_wdr \$PDDIR/plugin/_phcfwd/dic_wr \$PDDIR/plugin/_phcfwd/dic_wyd \$PDDIR/plugin/_phcfwd/dic_wydr \$PDDIR/plugin/_phcfwd/dic_wyr \$PDDIR/plugin/_phcfwd/dic_yd \$PDDIR/plugin/_phcfwd/dic_ydr \$PDDIR/plugin/_phcfwd/dic_yr | - | 同義語辞書格納ディレクトリ |
| \$PDDIR/plugin/_phcfwd/etc | - | FREEWORD プラグイン登録用ファイル格納ディレクトリ |

| ディレクトリ名 | ファイル名 | 説明 |
|-------------------------------|-------------|--|
| | _phcfwd.adt | FREEWORD プラグイン定義ファイル |
| | _phcfwd.pic | FREEWORD プラグイン PIC ファイル |
| \$PDDIR/plugin/_phcfwd/sample | - | FREEWORD プラグインのサンプルファイルを格納するディレクトリ |
| | bes01 | BES 単位指定コンフィグレーションファイル サンプルファイル |
| | pdplgsys | システム単位指定コンフィグレーションファイル サンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイル サンプルファイル |
| \$PDDIR/plugin/_phcfwd/spool | - | FREEWORD プラグインの障害情報格納ディレクトリ |
| \$PDDIR/plugin/_phcfwd/tmp | - | 作業用ファイル格納ディレクトリ |
| \$PDDIR/plugin/_phxfwr | - | IXFREEWORD プラグインを格納するディレクトリ |
| \$PDDIR/plugin/_phxfwr/bin | - | IXFREEWORD プラグインのユティリティを格納するディレクトリ |
| | phxidxls | インデクス情報の取得ユティリティ |
| | phxincmt | インデクスのインクリメンタルユティリティ |
| | phxexpnd | インデクスの拡張ユティリティ |
| | phxstartidx | インデクス作成開始ユティリティ |
| | phxcond | インデクス詰め替えユティリティ |
| | phxmodidx | インデクス情報変更ユティリティ |
| | phxdelay | インデクス分割遅延登録運用変更ユティリティ |
| \$PDDIR/plugin/_phxfwr/conf | - | IXFREEWORD プラグインのコンフィグレーションファイルを格納するディレクトリ |
| \$PDDIR/plugin/_phxfwr/etc | - | IXFREEWORD プラグイン登録用ファイル格納ディレクトリ |
| | _phxfwr.idx | IXFREEWORD プラグイン定義ファイル |
| | _phxfwr.pic | IXFREEWORD プラグイン PIC ファイル |
| \$PDDIR/plugin/_phxfwr/sample | - | IXFREEWORD プラグインのサンプルファイルを格納するディレクトリ |
| | pdplgsys | システム単位指定コンフィグレーションファイル サンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイル サンプルファイル |
| | bes01 | BES 単位指定コンフィグレーションファイル サンプルファイル |

| ディレクトリ名 | ファイル名 | 説明 |
|------------------------------|-------|-------------------------------|
| \$PDDIR/plugin/_phxfwr/spool | - | IXFREEWORD プラグインの障害情報格納ディレクトリ |
| \$PDDIR/plugin/_phxfwr/tmp | - | 作業用ファイル格納ディレクトリ |

(凡例)

- : 該当しない。

付録 B.2 ディレクトリおよびファイル構成 (Windows の場合)

ここでは、Text Search Plug-in を HiRDB に登録したあとの、SGML プラグインと n-gram インデクスプラグインのフォルダおよびファイル構成を表 B-3 に示します。また、FREEWORD プラグインと IXFREEWORD プラグインのフォルダおよびファイル構成を表 B-4 に示します。

表 B-3 SGML プラグインと n-gram インデクスプラグインのフォルダおよびファイル構成 (Windows の場合)

| フォルダ名 | ファイル名 | 説明 |
|-------------------------------|--------------------|-----------------------------|
| %PDDIR%\%plugin%_phsgml | - | SGML プラグインを格納するフォルダ |
| %PDDIR%\%plugin%_phsgml%\bin | - | SGML プラグインのユーティリティを格納するフォルダ |
| | phsregsetup.exe | レジストリ初期化ユーティリティ |
| | phssgm_lreg.exe | SGML 定義情報登録ユーティリティ |
| | phssgm_lrm.exe | SGML 定義情報削除ユーティリティ |
| | phssgm_lval.exe | SGML 定義情報参照ユーティリティ |
| | phssyn_dicreg.exe | 同義語辞書登録ユーティリティ |
| | phssyn_dicconv.bat | 同義語辞書移行ユーティリティ |
| | phssyn_dicrm.exe | 同義語辞書削除ユーティリティ |

| フォルダ名 | ファイル名 | 説明 |
|---|-------------|------------------------------------|
| %PDDIR%\plugin_phsgml\conf | - | SGML プラグインのコンフィグレーションファイルを格納するフォルダ |
| %PDDIR%\plugin_phsgml\dic %PDDIR%\plugin_phsgml\dic_c %PDDIR%\plugin_phsgml\dic_w %PDDIR%\plugin_phsgml\dic_y %PDDIR%\plugin_phsgml\dic_cw %PDDIR%\plugin_phsgml\dic_cy %PDDIR%\plugin_phsgml\dic_wy %PDDIR%\plugin_phsgml\dic_cwy %PDDIR%\plugin_phsgml\dic_cd %PDDIR%\plugin_phsgml\dic_cdr %PDDIR%\plugin_phsgml\dic_cr %PDDIR%\plugin_phsgml\dic_cwd %PDDIR%\plugin_phsgml\dic_cwdr %PDDIR%\plugin_phsgml\dic_cwr %PDDIR%\plugin_phsgml\dic_cwyd %PDDIR%\plugin_phsgml\dic_cwydr %PDDIR%\plugin_phsgml\dic_cwyr %PDDIR%\plugin_phsgml\dic_cydr %PDDIR%\plugin_phsgml\dic_cyr %PDDIR%\plugin_phsgml\dic_d %PDDIR%\plugin_phsgml\dic_dr %PDDIR%\plugin_phsgml\dic_r %PDDIR%\plugin_phsgml\dic_wd %PDDIR%\plugin_phsgml\dic_wdr %PDDIR%\plugin_phsgml\dic_wr %PDDIR%\plugin_phsgml\dic_wyd %PDDIR%\plugin_phsgml\dic_wydr %PDDIR%\plugin_phsgml\dic_wyr %PDDIR%\plugin_phsgml\dic_yd %PDDIR%\plugin_phsgml\dic_ydr %PDDIR%\plugin_phsgml\dic_yr | - | 同義語辞書格納フォルダ |
| %PDDIR%\plugin_phsgml\etc | - | SGML プラグイン登録用ファイル格納フォルダ |
| | _phsgml.adt | SGML プラグイン定義ファイル |
| | _phsgml.pic | SGML プラグイン PIC ファイル |
| %PDDIR%\plugin_phsgml\sample | - | SGML プラグインのサンプルファイルを格納するフォルダ |
| | bes01 | BES 単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgsys | システム単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイルサンプルファイル |
| %PDDIR%\plugin_phsgml\spool | - | SGML プラグインの障害情報格納フォルダ |

| フォルダ名 | ファイル名 | 説明 |
|------------------------------|------------------|---|
| %PDDIR%\plugin¥_phsgml¥tmp | - | 作業用ファイル格納フォルダ |
| %PDDIR%\plugin¥_phngram | - | n-gram インデクスプラグインを格納するフォルダ |
| %PDDIR%\plugin¥_phngram¥bin | - | n-gram インデクスプラグインのユティリティを格納するフォルダ |
| | phnidxls.exe | インデクス情報の取得ユティリティ |
| | phnincrement.exe | インデクスのインクリメンタルユティリティ |
| | phnmerge.exe | インデクスのマージユティリティ |
| | phnexpand.exe | インデクスの拡張ユティリティ |
| | phnstartidx.exe | インデクス作成開始ユティリティ |
| | phncond.exe | インデクス詰め替えユティリティ |
| | phnmodidx.exe | インデクス情報変更ユティリティ |
| | phndelay.exe | インデクス分割遅延登録運用変更ユティリティ |
| | phnconget.exe | 統計情報収集ユティリティ |
| | phnconbuild.exe | 統計情報統合ユティリティ |
| | phnconrstr.exe | 統計情報反映ユティリティ |
| | phnconchgcon.exe | 特徴ターム取得 RD エリア指定ユティリティ |
| %PDDIR%\plugin¥_phngram¥conf | - | n-gram インデクスプラグインのコンフィグレーションファイルを格納するフォルダ |
| %PDDIR%\plugin¥_phngram¥etc | - | n-gram インデクスプラグイン登録用ファイル格納フォルダ |
| | _phngram.idx | n-gram インデクスプラグイン定義ファイル |
| | _phngram.pic | n-gram インデクスプラグイン PIC ファイル |

| フォルダ名 | ファイル名 | 説明 |
|--------------------------------|------------|-------------------------------------|
| %PDDIR%\plugin¥_phngram¥sample | - | n-gram インデクスプラグインのサンプルファイルを格納するフォルダ |
| | pdplgsys | システム単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイルサンプルファイル |
| | bes01 | BES 単位指定コンフィグレーションファイルサンプルファイル |
| %PDDIR%\plugin¥_phngram¥spool | - | n-gram インデクスプラグインの障害情報格納フォルダ |
| %PDDIR%\plugin¥_phngram¥tmp | - | 作業用ファイル格納フォルダ |

(凡例)

- : 該当しない。

表 B-4 FREEWORD プラグインと IXFREEWORD プラグインのフォルダおよびファイル構成 (Windows の場合)

| フォルダ名 | ファイル名 | 説明 |
|-----------------------------|------------------|--|
| %PDDIR%\plugin¥_phcfwd | - | FREEWORD プラグインを格納するフォルダ |
| %PDDIR%\plugin¥_phcfwd¥bin | - | FREEWORD プラグインのユティリティを格納するフォルダ |
| | phcsyndicreg.exe | 同義語辞書登録ユティリティ |
| | phcsyndicrm.exe | 同義語辞書削除ユティリティ |
| | phcsyndiconv.bat | 同義語辞書移行ユティリティ |
| %PDDIR%\plugin¥_phcfwd¥conf | - | FREEWORD プラグインのコンフィグレーションファイルを格納するフォルダ |

| フォルダ名 | ファイル名 | 説明 |
|---|-------------|----------------------------------|
| %PDDIR%\plugin_phcfwd\dic %PDDIR%\plugin_phcfwd\dic_c %PDDIR%\plugin_phcfwd\dic_w %PDDIR%\plugin_phcfwd\dic_y %PDDIR%\plugin_phcfwd\dic_cw %PDDIR%\plugin_phcfwd\dic_cy %PDDIR%\plugin_phcfwd\dic_wy %PDDIR%\plugin_phcfwd\dic_cwy %PDDIR%\plugin_phcfwd\dic_cd %PDDIR%\plugin_phcfwd\dic_cdr %PDDIR%\plugin_phcfwd\dic_cr %PDDIR%\plugin_phcfwd\dic_cwd %PDDIR%\plugin_phcfwd\dic_cwdr %PDDIR%\plugin_phcfwd\dic_cwr %PDDIR%\plugin_phcfwd\dic_cwyd %PDDIR%\plugin_phcfwd\dic_cwydr %PDDIR%\plugin_phcfwd\dic_cwyr %PDDIR%\plugin_phcfwd\dic_cyd %PDDIR%\plugin_phcfwd\dic_cydr %PDDIR%\plugin_phcfwd\dic_cyr %PDDIR%\plugin_phcfwd\dic_d %PDDIR%\plugin_phcfwd\dic_dr %PDDIR%\plugin_phcfwd\dic_r %PDDIR%\plugin_phcfwd\dic_wd %PDDIR%\plugin_phcfwd\dic_wdr %PDDIR%\plugin_phcfwd\dic_wr %PDDIR%\plugin_phcfwd\dic_wyd %PDDIR%\plugin_phcfwd\dic_wydr %PDDIR%\plugin_phcfwd\dic_wyr %PDDIR%\plugin_phcfwd\dic_yd %PDDIR%\plugin_phcfwd\dic_ydr %PDDIR%\plugin_phcfwd\dic_yr | - | 同義語辞書格納フォルダ |
| %PDDIR%\plugin_phcfwd\etc | - | FREEWORD プラグイン登録用ファイル格納フォルダ |
| | _phcfwd.adt | FREEWORD プラグイン定義ファイル |
| | _phcfwd.pic | FREEWORD プラグイン PIC ファイル |
| %PDDIR%\plugin_phcfwd\sample | - | FREEWORD プラグインのサンプルファイルを格納するフォルダ |
| | bes01 | BES 単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgsys | システム単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイルサンプルファイル |
| %PDDIR%\plugin_phcfwd\spool | - | FREEWORD プラグインの障害情報格納 / フォルダ |
| %PDDIR%\plugin_phcfwd\tmp | - | 作業用ファイル格納フォルダ |

| フォルダ名 | ファイル名 | 説明 |
|--------------------------------|-----------------|--|
| %PDDIR%\plugin¥_phxfwrд | - | IXFREEWORD プラグインを格納するフォルダ |
| %PDDIR%\plugin¥_phxfwrд¥bin | - | IXFREEWORD プラグインのユーティリティを格納するフォルダ |
| | phxidxls.exe | インデクス情報の取得ユーティリティ |
| | phxincmt.exe | インデクスのインクリメンタルユーティリティ |
| | phxexpnd.exe | インデクスの拡張ユーティリティ |
| | phxstartidx.exe | インデクス作成開始ユーティリティ |
| | phxcond.exe | インデクス詰め替えユーティリティ |
| | phxmodidx.exe | インデクス情報変更ユーティリティ |
| | phxdelay.exe | インデクス分割遅延登録運用変更ユーティリティ |
| %PDDIR%\plugin¥_phxfwrд¥conf | - | IXFREEWORD プラグインのコンフィグレーションファイルを格納するフォルダ |
| %PDDIR%\plugin¥_phxfwrд¥etc | - | IXFREEWORD プラグイン登録用ファイル格納フォルダ |
| | _phxfwrд.idx | IXFREEWORD プラグイン定義ファイル |
| | _phxfwrд.pic | IXFREEWORD プラグイン PIC ファイル |
| %PDDIR%\plugin¥_phxfwrд¥sample | - | IXFREEWORD プラグインのサンプルファイルを格納するフォルダ |
| | pdplgsys | システム単位指定コンフィグレーションファイルサンプルファイル |
| | pdplgutsys | ユニット単位指定コンフィグレーションファイルサンプルファイル |
| | bes01 | BES 単位指定コンフィグレーションファイルサンプルファイル |
| %PDDIR%\plugin¥_phxfwrд¥spool | - | IXFREEWORD プラグインの障害情報格納フォルダ |
| %PDDIR%\plugin¥_phxfwrд¥tmp | - | 作業用ファイル格納フォルダ |

(凡例)

- : 該当しない。

付録 C 資源の排他

Text Search Plug-in は、HiRDB の資源である論理ファイルを使用する場合があります、HiRDB の排他制御の規則に従っています。

ここでは、Text Search Plug-in が使用する資源の排他規則について説明します。

付録 C.1 排他取得内容

(1) SGML プラグインおよび n-gram インデクスプラグインの場合

SGML プラグインおよび n-gram インデクスプラグインの論理ファイルに対する排他制御のモードを表 C-1 および表 C-2 に示します。論理ファイルと HiRDB の排他制御の詳細については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

表 C-1 SQL 実行時の排他制御のモード (SQL に対応した論理ファイルの使用)

| SQL 文 | SGML プラグイン | | n-gram インデクスプラグイン | |
|---------|------------|--------|-------------------|--------|
| | 同時実行なし | 同時実行あり | 同時実行なし | 同時実行あり |
| 検索 | PR | SR | PR | SR |
| 更新 | EX | SU | EX | PU |
| 追加 | EX | SU | EX | PU |
| 削除 | × | × | EX | PU |
| LOCK 文 | × | × | × | × |
| 表削除 | × | × | × | × |
| インデクス定義 | × | × | EX | EX |
| インデクス削除 | × | × | × | × |
| 全行削除 | × | × | EX | PU |
| 表定義変更 | × | × | × | × |

(凡例)

- PR：共用モード
- EX：排他モード
- SR：意図共用モード
- SU：意図排他モード
- PU：共用意図排他モード
- ×：同時実行できません。

注

データ更新時は、Text Search Plug-in のデータの整合性確保のために一時的に排他モード (EX) へ遷移しますが、整合性確保後、遷移前の排他制御のモードへ戻ります。

表 C-2 ユティリティコマンドの排他制御のモード (SGML プラグインおよび n-gram インデクスプラグイン)

| コマンド | 表格 納 RD エリ ア | 抽象 データ 型 列定義 LOB 用 RD エリア | 表 | ノーウ エイト 表 | インデ クス格 納 RD エリア | イン デク ス | ページ | 行 | キー 値 | 論理 ファ イル |
|-------------|--------------------------|---|----|-----------------|---------------------------|---------------|-----|---|---------|----------------|
| phnidxls | - | - | - | - | ex/sr | - | - | - | - | ex/sr |
| phnincmt | - | - | - | - | ex | - | - | - | - | ex |
| phnexpnd | - | - | - | - | ex | - | - | - | - | ex |
| phnmerge | SR | SR | PR | - | SU | EX/ PU | - | - | - | EX/ PU |
| phnmodidx | - | - | - | - | ex | - | - | - | - | ex |
| phnstartidx | SR | SR | PR | - | SU | EX/ PU | - | - | - | EX/ PU |
| phncond | - | - | - | - | ex | - | - | - | - | ex |
| phnconget | SR | SR | PR | - | SU | EX/ PU | - | - | - | PR |
| phnconbuild | - | - | - | - | - | - | - | - | - | - |
| phnconrstr | - | - | - | - | SU | EX | - | - | - | EX |
| phnchgcon | - | - | - | - | SU | EX | - | - | - | EX |
| phndelay | SR | SR | PR | - | EX | EX | - | - | - | EX |

(凡例)

左 / 右 : 左 = 同時実行なし 右 = 同時実行あり

PR または pr : 共用モード

EX または ex : 排他モード

SR または sr : 意図共用モード

SU または su : 意図排他モード

PU または pu : 共用意図排他モード

- : 該当しません。

大文字は、表またはインデクスが存在するすべてのサーバ上で排他を取得します。

小文字は、指定された RD エリアが存在するサーバ上でだけ排他を取得します。

(2) FREEWORD プラグインおよび IXFREEWORD プラグインの場合

FREEWORD プラグインおよび IXFREEWORD プラグインの論理ファイルに対する排他制御のモードを表 C-3 および表 C-4 に示します。論理ファイルと HiRDB の排他制御の詳細については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

表 C-3 SQL 実行時の排他制御のモード (SQL に対応した論理ファイルの使用)

| SQL 文 | FREEWORD プラグイン | | IXFREEWORD プラグイン | |
|---------|----------------|--------|------------------|--------|
| | 同時実行なし | 同時実行あり | 同時実行なし | 同時実行あり |
| 検索 | - | - | PR | SR |
| 更新 | - | - | EX | PU |
| 追加 | - | - | EX | PU |
| 削除 | - | - | EX | PU |
| LOCK 文 | - | - | - | - |
| 表削除 | - | - | - | - |
| インデクス定義 | - | - | EX | EX |
| インデクス削除 | - | - | - | - |
| 全行削除 | - | - | EX | PU |
| 表定義変更 | - | - | - | - |

(凡例)

PR : 共用モード

EX : 排他モード

SR : 意図共用モード

PU : 共用意図排他モード

- : 該当しません。

注

データ更新時は、Text Search Plug-in のデータの整合性確保のために一時的に排他モード (EX) へ遷移しますが、整合性確保後、遷移前の排他制御のモードへ戻ります。

表 C-4 ユティリティコマンドの排他制御のモード

| コマンド | 表格納 RD エリア | 抽象データ型列定義 LOB 用 RD エリア | 表 | ノーウエイト表 | インデクス格納 RD エリア | インデクス | ページ | 行 | キー値 | 論理ファイル |
|-------------|------------|------------------------|----|---------|----------------|-------|-----|---|-----|--------|
| phxidxls | - | - | - | - | ex/sr | - | - | - | - | ex/sr |
| phxinemt | - | - | - | - | ex | - | - | - | - | ex |
| phxexpnd | - | - | - | - | ex | - | - | - | - | ex |
| phxstartidx | SR | SR | PR | - | SU | EX | - | - | - | EX/PU |
| phxcond | - | - | - | - | ex | - | - | - | - | ex |

(凡例)

左 / 右 : 左 = 同時実行なし 右 = 同時実行あり
PR または pr : 共用モード
EX または ex : 排他モード
SR または sr : 意図共用モード
SU または su : 意図排他モード
PU または pu : 共用意図排他モード
- : 該当しません。

大文字は、表またはインデクスが存在するすべてのサーバ上で排他を取得します。
小文字は、指定された RD エリアが存在するサーバ上でだけ排他を取得します。

付録 C.2 レジストリの排他

SGMLTEXT 型列を含む表の行に対して、UAP 実行中に、「phsregsetup (レジストリ初期化ユティリティ)」、「phssgmlreg (SGML 定義情報登録ユティリティ)」、「phssgmlrm (SGML 定義情報削除ユティリティ)」または「phssgmlval (SGML 定義情報参照ユティリティ)」を実行するとエラーになる場合がありますのでご注意ください。

付録 C.3 デッドロックの防止

SGMLTEXT 型列を含む表の行に対して、同一トランザクション内で行削除 (DELETE) に続いてほかの処理を実行した場合、デッドロックが発生するおそれがあります。この場合は、行削除 (DELETE) のあとにコミット (COMMIT) を実行してください。

付録 D 障害発生時の運用方法

Text Search Plug-in で障害が発生した場合、HiRDB の管理者は出力されたメッセージを確認し、「付録 F エラーメッセージ」を参照して、障害の原因を調べてください。

出力されたメッセージが HiRDB 本体からの場合、必要な処置は HiRDB の障害発生時の運用に従います。マニュアル「HiRDB メッセージ」または「HiRDB システム運用ガイド」を参照して対処してください。

なお、Text Search Plug-in 固有の運用について次に示します。

バックアップからの全面的な回復操作では、同義語辞書は回復されません。この場合、同義語辞書テキストファイルから、同義語辞書を再作成してください。

Text Search Plug-in 実行時に障害が発生した場合は、障害情報を出力します。障害情報が不要となった場合は、次のディレクトリ内のファイルをすべて削除してください。ファイルの削除は、HiRDB を停止させたあとで実行してください。

- UNIX の場合

```
$PDDIR/plugin/_phsgml/spool
$PDDIR/plugin/_phsgml/tmp
$PDDIR/plugin/_phngram/spool
$PDDIR/plugin/_phngram/tmp
$PDDIR/plugin/_phcfwd/spool
$PDDIR/plugin/_phcfwd/tmp
$PDDIR/plugin/_phxfwd/spool
$PDDIR/plugin/_phxfwd/tmp
```

- Windows の場合

```
%PDDIR%\plugin\_phsgml\spool
%PDDIR%\plugin\_phsgml\tmp
%PDDIR%\plugin\_phngram\spool
%PDDIR%\plugin\_phngram\tmp
%PDDIR%\plugin\_phcfwd\spool
%PDDIR%\plugin\_phcfwd\tmp
%PDDIR%\plugin\_phxfwd\spool
%PDDIR%\plugin\_phxfwd\tmp
```

付録 E 系切り替え

ここでは、Text Search Plug-in の系切り替え運用について説明します。

HiRDB の系切り替え運用を行う場合は、HiRDB の各種設定に加えて Text Search Plug-in として以下の設定が必要です。

(1) 同義語辞書

同義語辞書は待機系にも設定する必要があります。辞書の登録は次のどちらかの方法で行なってください。

- 待機系 HiRDB を起動した状態で、`phs(c)syndicreg` コマンドを使って待機系に同義語辞書を登録する。
- 現用系に登録されている以下の同義語辞書データを待機系にコピーする。

UNIX の場合

- SGMLTEXT 型
`$PDDIR/plugin/_phnggml/dic`
- FREEWORD 型の場合
`$PDDIR/plugin/_phcfwd/dic`

Windows の場合

- SGMLTEXT 型
`%PDDIR%\plugin¥_phnggml¥dic`
- FREEWORD 型の場合
`%PDDIR%\plugin¥_phcfwd¥dic`

注

ディレクトリ名称が `dic` で始まるすべてのディレクトリ、およびディレクトリ内ファイルを待機系にコピーします。

(2) インデクス定義ファイル

次のディレクトリにインデクス定義ファイルが存在する場合は、待機系にコピーします。

UNIX の場合

- NGRAM 型
`$PDDIR/plugin/_phngram/idx_def`
- IXEFREEWORD 型の場合
`$PDDIR/plugin/_phxfwd/idx_def`
- 英文デリミタファイル
- 英文削除コードファイル
- 英文不要語ファイル

Windows の場合

- NGRAM 型
%PDDIR%¥plugin¥_phngram¥idx_def
- IXEFREEWORD 型の場合
%PDDIR%¥plugin¥_phxfwr¥idx_def
- 英文デリミタファイル
- 英文削除コードファイル
- 英文不要語ファイル

(3) 環境定義ファイル

現用系と同じ定義で運用してください。

ただし、メモリ・排他資源数等のマシン環境に依存する部分については、待機系マシンに合わせた設定としてください。

付録 F エラーメッセージ

ここでは、Text Search Plug-in のエラーメッセージについて説明します。

Text Search Plug-in でエラーが発生した場合、ユーザエラーのメッセージは SQL 連絡領域に、システムで発生したエラーのメッセージはメッセージログに、Text Search Plug-in のユティリティのメッセージは標準エラー出力にそれぞれ出力されます。

付録 F.1 SQL 連絡領域に出力されるメッセージ

(1) エラーコードの詳細化指定をしない場合

エラーコードの詳細化指定をしない場合、SQL 連絡領域にメッセージが返却されるときは、HiRDB のメッセージ KFPA11997-E とともに出力されます。出力される形式は次のとおりです。

KFPA11997-E Error occurred during execution of PLUGIN function, "aa....aa"

aa....aa に Text Search Plug-in が出力したメッセージが設定されます。

KFPA11997-E のメッセージの詳細については、マニュアル「HiRDB メッセージ」を参照してください。

(2) エラーコードの詳細化指定をする場合

エラーコードの詳細化指定をする場合、SQL 連絡領域にメッセージが返却されるときは、HiRDB のメッセージ KFPA18xxx-E とともに出力されます。出力される形式は次のとおりです。なお、エラーコードの詳細化指定をするかどうかは、環境定義ファイルで指定できます。

KFPA18xxx-E (dddddd)aa....aa

dddddd : プラグインモジュール名

Text Search Plug-in では次のどれかが出力されます。

_phsgml : SGML プラグインで出力したメッセージ

_phngram : n-gram インデクスプラグインで出力したメッセージ

_phcfwd : FREEWORD プラグインで出力したメッセージ

_phxfwr : IXFREEWORD プラグインで出力したメッセージ

aa....aa に Text Search Plug-in が出力したメッセージが設定されます。

xxx は Text Search Plug-in が出力した SQL コードの下 3 けたが設定されます。出力さ

れる SQL コードを次に示します。

表 F-1 SQL コード一覧

| 分類 | 詳細要因 | 対処 | SQL コード |
|---------------------|------------------------------------|--|---------|
| リトライ可能エラー | 排他エラー, またはロールバック中に実行 | 再実行してください。 | -3000 |
| | メモリ不足 | <ul style="list-style-type: none"> 再実行してください。 システムリソースを見積もり直して再実行してください。 | -3001 |
| リソース関連エラー | 共用メモリ不足 | リソース定義を見直してください。 | -3020 |
| | RD エリア容量不足 | リソースを見積もり直してください。 | -3021 |
| | その他リソース不足 | リソースを見積もり直して再実行してください。 | -3022 |
| 登録データエラー | データ不正 | 登録データを見直してください。 | -3040 |
| SQL エラー | 検索条件不正 (文法エラー) | 検索条件を見直してください。 | -3061 |
| | CREATE TABLE / CREATE INDEX 定義情報不正 | PLUGIN 句を見直してください。 | -3062 |
| | 各種制限オーバー | <ul style="list-style-type: none"> 定義ファイルを見直してください。 検索条件を見直してください。 | -3063 |
| | 検索条件不正 / 実行機能不正 | <ul style="list-style-type: none"> 検索条件を見直してください。 インデクスを見直してください。 | -3064 |
| 運用修正 | 運用定義不正 | 運用定義を見直してください。 | -3080 |
| | 各種定義ファイル内容不正, またはインデクスが存在しない | <ul style="list-style-type: none"> 各定義ファイルを見直して再実行してください。 インストール環境を見直してください。 運用環境を見直してください。 | -3081 |
| | 関連プログラム不足 | 関連プログラムを見直してください。 | -3082 |
| 環境不正 (ファイル指定, ディスク) | OS ファイル I / O エラー | ファイルおよびディスクの確認をおこなってから, 必要に応じて回復作業をおこなってください。 | -3100 |
| | DB ファイル I / O エラー | ディスクを確認してから, 必要に応じて回復作業をおこなってください。 | -3101 |

| 分類 | 詳細要因 | 対処 | SQL コード |
|-----|---------------------|--|---------|
| その他 | HiRDB エラー | <ul style="list-style-type: none"> HiRDB のメッセージに従って対処してください。 保守員に連絡してください。 | -3200 |
| | サーバエラー | システム管理者に連絡してください。 | -3201 |
| | HiRDB セットアップ文字コード不正 | システム管理者に連絡してください。 | -3202 |
| | システムエラー | 内部矛盾が発生しました。保守員に連絡してください。 | -3299 |

付録 F.2 メッセージログ・標準エラー出力に出力されるメッセージ

メッセージログ・標準エラー出力に出力される場合、HiRDB のメッセージ KFPY99999-I とともに出力されます。出力される形式は次のとおりです。

```
KFPY99999-I (dddddd)xxxxx-y : zzzzz
```

dddddd : プラグインモジュール名

Text Search Plug-in では下記のどれかが出力されます。

- _phsgml : SGML プラグインで出力したメッセージ
- _phngram : n-gram インデクスプラグインで出力したメッセージ
- _phcfwd : FREEWORD プラグインで出力したメッセージ
- _phxfwd : IXFREEWORD プラグインで出力したメッセージ

xxxxx : Text Search Plug-in でのメッセージ ID

y : エラーメッセージの重要度

次のどれかが出力されます。

- I : 情報メッセージ
- W : 警告メッセージ
- E : エラーメッセージ
- Q : 応答待ちメッセージ

zzzzz : Text Search Plug-in の出力したメッセージ

メッセージログに出力されたメッセージは、HiRDB の運用コマンド「pdcatt (ファイルの内容表示)」で標準エラー出力に出力できます。

付録 F.3 メッセージの記述形式

このマニュアルで説明するメッセージの記述形式を次に示します。

(1) SQL 連絡領域に出力されるメッセージ

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。

(P) メッセージを受け取ったプログラムの処置を示します。

[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。

(2) メッセージログ・標準エラー出力に出力されるメッセージ

XXXXX-Y 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。

(P) メッセージを受け取ったプログラムの処置を示します。

[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。

付録 F.4 メッセージ一覧

(1) SQL 連絡領域に出力されるメッセージ

(A)

Asterisk symbol exists between pipe symbol and term.

前方一致,後方一致指定の"|" (パイプ) とターム条件の間に "*" (ワイルドカード) が指定されています。

(P) 検索条件式を修正して再実行してください。

Attribute exist in specific_and. offset=aa

特定構造検索指定に属性指定があります。

aa: 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

(B)

Brace symbol not match in condition.

検索条件式中の波括弧 ({ , }) が対応していません。

(P) 検索条件式を修正して再実行してください。

Bracket symbol not match in condition.

検索条件式中の角括弧 ([,]) が対応していません。
 (P) 検索条件式を修正して再実行してください。

(C)

Cannot open kanarule file. errno=aa file=bb

カタカナルールファイルのオープンに失敗しました。
 aa : エラー番号
 bb : カタカナルールファイル名
 [対策] 保守員に連絡してください。

Cannot read kanarule file. errno=aa file=bb

カタカナルールファイルの読み込みに失敗しました。
 aa : エラー番号
 bb : カタカナルールファイル名
 [対策] 保守員に連絡してください。

Cannot seek kanarule file. errno=aa file=bb

カタカナルールファイルのシークに失敗しました。
 aa : エラー番号
 bb : カタカナルールファイル名
 [対策] 保守員に連絡してください。

Cannot specify "not" in concept condition. offset=aa

概念検索では検索因子に NOT を指定できません。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

Cannot specify "proximity" in concept condition. offset=aa

概念検索では近傍検索条件を指定できません。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

Cannot specify search factor "and"/"or" in concept condition. offset=aa

概念検索では検索因子を AND/OR で接続できません。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

Cannot specify "specific_and" in concept condition. offset=aa

概念検索では specific_and を指定できません。

aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Cannot specify "specific_and" in freeword condition. offset=aa

文字列検索では specific_and を指定できません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Cannot specify struct in freeword condition. offset=aa

文字列検索では構造名を指定できません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Cannot specify term of one character in proximity.

ワイルドカード検索または近傍検索に 1 文字の検索条件が含まれています。
(P) 検索条件を変更して再実行してください。

Column not identified.

表・列が特定できません。
(P) 抽象データ型関数の列指定に SGMLTEXT 型列が正しく指定されているか確認してください。SGMLTEXT 型列を指定して再実行してください。

Comma not exist. offset=aa

検索条件式中の指定が必要な個所に "," (コンマ) が見つかりません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Comma symbol not exist in condition.

検索条件式中の指定が必要な個所に "," (コンマ) が見つかりません。
(P) 検索条件式を修正して再実行してください。

Config File I/O error.filename=aa..aa,func=bb..bb,code=XX,file=cc..cc,line=YY

コンフィグレーションファイルのアクセス中にエラーを検出しました。
aa..aa : ファイル名
bb..bb : オペレーション
XX : 詳細情報
cc..cc : オペレーションを発行したソースファイル名
YY : オペレーションを発行した行
(P) HiRDB 管理者に連絡してください。
[対策] メッセージログの内容を確認し、メッセージが出力されている場合はそれに従っ

て対処してください。出力されていない場合は、保守員に連絡してください。

(D)

Define file not exist.filename= aa..aa,func=bb..bb,file=cc..cc,line=XX

指定されたインデクスデリミタファイル、不要語ファイル、または削除コードファイルが存在しません。または、指定できる個数の上限を超えています。

aa..aa：ファイル名

bb..bb：オペレーション

cc..cc：エラーの発生したファイル

XX：エラーの発生した行

[対策] プラグインオプションに指定したファイル名称を確認してください。

Directory or File for Concept Information not exist or destroyed.

概念情報として指定されたディレクトリまたはファイルが存在しないか、または破壊されています。

(P) 指定したディレクトリまたはファイルを確認してください。

Document size is 0 bytes.

登録しようとしている文書の文書サイズが、0 バイトです。

(P) 文書サイズが 0 バイトの文書を登録する場合は、NULL 値を登録してください。

Document size is too large.

登録しようとした文書の文書サイズが、Text Search Plug-in の制限値を超えています。

(P) 登録文書の文書サイズを Text Search Plug-in の制限値以下にしてください。

Double quote symbol not exist in condition.

単純文字列指定の前後に必要な " (ダブルクォーテーション) が見つかりません。

(P) 検索条件式を修正して再実行してください。

DTD name for display is not specified in extract parameter.

extracts パラメタに表示用 DTD 名の指定がありません。

(P) HiRDB 管理者に連絡してください。

[対策] extracts パラメタの内容を確認してください。

DTD name is not specified in plugin description.

CREATE TABLE で、SGMLTEXT 型の列のプラグインオプションに DTD 名が指定されていません。

(P) プラグインオプションに、タグ <DTD></DTD> によって DTD 名を指定してください。

DTD name unmatched.'aa..aa' in sgml document, 'bb..bb' in plugin description.

登録しようとした SGML 文書中の DTD 名と、登録対象列のプラグインオプション中に定義した DTD 名とが一致しません。

aa..aa : SGML 文書中の DTD 名

bb..bb : 列定義の DTD 名

(P) 登録対象の SGML 文書が正しいか確認してください。または、登録対象列の定義内容を確認してください。

DTD(aa..aa) not catalogued in system.

CREATE TABLE の SGMLTEXT 型の列のプラグインオプションに指定した DTD がレジストりに登録されていません。

aa..aa : DTD 名

[対策] phssgmlreg (sgml 定義情報登録ユーティリティ) で必要な DTD を登録してください。

(E)

Each definition unmatched. value1=Index Option aa, value2=bb

aa と bb の定義の組み合わせは、実行できない組み合わせです。

aa : CREATE INDEX のプラグインオプションに指定したインデクス作成方法

bb : 環境定義またはユーティリティで指定した同時実行指定 ("Lockmode")

(P) aa と bb の定義内容を見直してください。

aa="CREATEMODE=3", bb="Lockmode" の場合、データ型プラグインの環境定義ファイルの phs_search_wait または phc_search_wait に「nowait_with_update」が指定されているかどうかを確認してください。

また、インデクス作成開始ユーティリティ (phnstartidx または phxstartidx) の -w オプションに、「nowait_with_update」が指定されているかどうかを確認してください。

Enable to use n-gram index.

全文検索対象に指定した SGMLTEXT 型、または FREEWORD 型の列に、インデクスが定義されていないため、全文検索できません。

または、発行した抽象データ型関数がインデクス専用関数のため、インデクスを使用した検索ができません。

(P) インデクスを定義していない場合は、その列に対して CREATE INDEX でインデクスを定義してください。

または、SQL 文を見直して再実行してください。なお、インデクス型プラグイン専用関数の制限については、マニュアル「HiRDB UAP 開発ガイド」を参照してください。

English option extend term library error occurred.(aa,bb)

aa..aa : 詳細情報

bb..bb : エラーの発生したファイル名

Text Search Enhancer for English のライブラリでエラーが発生しました。
 [対策] 保守員に連絡してください。

Equal not exist. offset=aa

検索条件式中の指定が必要な個所に = (イコール) がありません。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

Error exists in extracting structure name. ("," or ";")

抽出構造名の "," (コンマ) または ";" (セミコロン) の指定に誤りがあります。
 (P) 抽出構造名の指定内容を確認してください。内容を確認後、再実行してください。

Error in Plugin Option aa(XX)

プラグインオプションの指定に誤りがあります。
 aa : エラーの発生したファイル
 XX : エラーの発生した行
 [対策] プラグインオプションを修正し、再度実行してください。

Error occurred in Text Search Plug-in.reason=aa..aa

Text Search Plug-in でエラーが発生しました。
 aa..aa : エラー要因
 MEMORY : サーバでメモリ不足が発生しています。
 HiRDB : HiRDB のサービスでエラーが発生しています。
 TS Plug-in : Text Search Plug-in でエラーが発生しています。
 (P) エラー要因が MEMORY の場合は、しばらく待って再実行してください。それでも、エラーが発生する場合は HiRDB 管理者に連絡してください。
 エラー要因が MEMORY 以外の場合は、HiRDB 管理者に連絡してください。
 [対策] エラー要因が MEMORY の場合は、システムのメモリ状態を確認してください。そのほかの場合は、システムログメッセージの内容に従って対処してください。

Extract parameter(aa..aa) not catalogued in system.

CREATE TABLE のプラグインオプションに指定した extracts パラメタがレジストリに登録されていません。
 aa..aa : extracts パラメタ名
 (P) HiRDB 管理者に連絡してください。
 [対策] phssgmlreg (sgml 定義情報登録ユーティリティ) で必要な extracts パラメタを登録してください。

(F)

Fail to lock aa..aa.

aa..aa の排他に失敗しました。n-gram インデクス用 RD エリアまたは SGMLTEXT データ用 RD エリアへのアクセスが競合しています。

aa..aa : Index File (n-gram インデクス用 RD エリア)

SGML Logical File (SGMLTEXT データ用 RD エリア)

(P) しばらく待って再実行してください。

[対策]

- デッドロックが発生している場合
ユーザプログラムの資源排他を見直してください。
- 資源排他待ちによるタイムアウトが発生している場合
システム定義, ユニット定義に指定した排他待ち限界経過時間 (pd_lck_wait_time) を見直してください。

Fail to lock because file is over the process limit. filename=aa..aa, line=bb..bb.

1 プロセスのオープンファイル数が上限を超えました。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行

(P) しばらく待って再実行してください。

[対策] 再実行の際にも同様のエラーが発生した場合には、プロセスごとに対応してください。

- HiRDB のサーバプロセスの場合
HiRDB 管理者に連絡してください。
- ユーザ起動によるプロセス (コマンド・ユティリティ) の場合
OS のパラメタの maxfiles の設定値を変更してください。

Fail to lock because file is over the system limit. filename=aa..aa, line=bb..bb.

システム全体のオープンファイル数が上限を超えました。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行

(P) しばらく待って再実行してください。

[対策] 再実行の際にも同様のエラーが発生した場合には、次に示すように対応してください。

- OS のパラメタの nfile の設定値を変更する。
- コマンドを実行したサーバマシンで動作している不要なプロセスを停止する。
- コマンドを実行しない。

Fail to lock because of time over. file=aa..aa, line=bb..bb.

資源排他待ちによるタイムアウトが発生しました。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行
 (P) しばらく待って再実行してください。
 [対策] システム定義, ユニット定義に指定した排他待ち限界経過時間
 (pd_lck_wait_timeout) を見直してください。

Fail to thread lock. filename=aa..aa, line=bb..bb.

排他に失敗しました。
 aa..aa : ファイル名
 bb..bb : オペレーションを発行した行
 (P) しばらく待って再実行してください。

Fail to lock .file=aa..aa line=bb..bb.

排他に失敗しました。
 aa..aa : ファイル名
 bb..bb : オペレーションを発行した行
 (P) しばらく待って再実行してください
 [対策] このメッセージの前後に KFPS00440-I, KFPS00441-I, KFPS00451-I メッセージが出力されている場合はそれにしたがって対処してください。

Fatal error occurred. detail= aa.aa,file= bb..bb ,line= XX

システム内部で自己矛盾を検出しました。
 aa..aa : 詳細情報
 bb..bb : エラーの発生したファイル名
 XX : エラーの発生した行
 (P) 保守員に連絡してください。

File I/O error. filename=aa..aa ,func=bb..bb ,code=XX ,file= cc..cc ,line= YY

システムファイルのアクセス中にエラーを検出しました。
 aa..aa : ファイル名
 bb..bb : オペレーション
 XX : 詳細情報
 cc..cc : オペレーションを発行したソースファイル名
 YY : オペレーションを発行した行
 (P) HiRDB 管理者に連絡してください。
 [対策] メッセージログの内容を確認し, メッセージが出力されている場合はそれに従って対処してください。出力されていない場合は, 保守員に連絡してください。

File open error. filename = aa..aa , code =XX

aa..aa に示されるファイルのオープンに失敗しました。
 aa..aa : オープンできなかったファイル名
 XX : 詳細情報

(P) HiRDB 管理者に連絡してください。

[対策] メッセージログの内容を確認し、それに従って対処してください。出力されていない場合は、保守員に連絡してください。

File size is over the system limit. filename = aa..aa

aa..aa に示されるファイル容量がシステムの上限を超えました。

aa..aa : サイズがシステムの上限を超えたファイル名

(P) HiRDB 管理者に連絡してください。

[対策] ファイルサイズがシステムの上限を超えないように、インデクスを作成してください。

(1)

Illegal character exist after AND/OR. offset=aa

AND/OR 指定の後ろの文字が誤っています。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Illegal character exist after structure name.

構造名指定のあとに指定できない文字列があります。指定できる文字は、丸括弧 (

(,)), 波括弧 ({ , }), 角括弧 ([,]) です。

(P) 検索条件式を修正して再実行してください。

Illegal characters exist after back slash symbol.

"¥" (エスケープ文字) のあとに、指定できる文字以外を指定しています。

(P) 検索条件式を修正して再実行してください。

Illegal characters exist after back slash symbol in search term. offset=aa

検索ターム中の "¥" (エスケープ文字) の後ろの文字が誤っています。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Illegal characters exist after back slash symbol. offset=aa

"¥" (エスケープ文字) の後ろの文字が誤っています。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Illegal characters exist after condition. offset=aa

検索条件式の後ろに誤った文字があります。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Illegal characters exist in term condition.

ターム条件中に誤った文字があります。
(P) 検索条件式を修正して再実行してください。

Illegal expand option.

SYNONYM, SOUNDEX_EXP に指定している展開モード指定に, "A", "E", "K" 以外が指定されています。
(P) 展開モードの指定内容を修正して再実行してください。

Illegal expand option. offset=aa

異表記展開モード指定が誤っています。
aa: 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Illegal expansion function invocation.

展開関数に, SOUNDEX_EXP, SYNONYM 以外が指定されています。
(P) 検索条件式を修正して再実行してください。

Illegal number in proximity distance.

近傍条件に指定された距離の指定が指定可能範囲 (0 ~ 2,147,483,646) ではありません。
(P) 距離の指定を指定可能範囲 (0 ~ 2,147,483,646) にして, 再実行してください。

Illegal operand in proximity distance.

近傍距離指定の演算子に, >=, =, <= 以外が指定されています。
(P) 近傍条件の内容を修正して再実行してください。

Illegal operand in proximity distance. offset=aa

近傍距離指定の演算子に, >=, =, <= 以外が指定されています。
aa: 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Illegal order in proximity.

近傍条件の順序指定に, ANY_ORDER, IN_ORDER 以外が指定されています。
(P) 近傍条件の順序指定の内容を修正して再実行してください。

Illegal order in proximity. offset=aa

近傍条件の順序指定が誤っています。
aa: 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Illegal position of pipe symbol.

前方一致、後方一致指定の "|" (パイプ) の位置が誤っています。
(P) 前方一致、後方一致指定の場合は、単純文字列の最前部または最後尾に指定してください。
単なる文字として指定する場合は直前に "¥" (エスケープ文字) を付けてください。

Illegal search factor exist. offset=aa

検索因子が誤っています。
aa: 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Illegal structure weight.weight=XX

構造の重みの値が指定可能範囲 (1 ~ 100) ではありません。
XX: 指定された重みの値
(P) 構造の重みの値を指定可能範囲 (1 ~ 100) にして再実行してください。

Illegal symbol exist in search term. ("*" or "?")

検索タームの "*" または "?" の指定に誤りがあります。
[対策] 検索条件式を修正して再実行してください。

Illegal term weight. Weight=XX

タームの重みの値が指定可能範囲 (1 ~ 100) ではありません。
XX: 指定された重みの値
(P) タームの重みの値を指定可能範囲 (1 ~ 100) にして再実行してください。

Illegal text function invocation.

展開指定 (SOUNDEX_EXP, SYNONYM), 近傍条件 (PROXIMITY) または単純文字列指定が必要な所にこれらの指定がありません。
(P) 中括弧のあとに、展開指定 (SOUNDEX_EXP, SYNONYM), 近傍条件 (PROXIMITY) または単純文字列指定があるか確認してください。その後、検索条件式を修正して再実行してください。

Illegal unit in proximity.

近傍条件の距離単位指定に、CHARACTERS 以外が指定されています。
(P) 近傍条件の距離単位指定に、CHARACTERS を指定して再実行してください。
英文検索の場合
英文検索で近傍条件の距離単位指定に、WORDS 以外が指定されています。
[対策] 英文検索の場合は、近傍条件の距離単位指定に、WORDS を指定して再実行してください。

Illegal unit in proximity. offset=aa

近傍条件の距離単位が誤っています。

aa：検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Incomplete term condition.

ターム条件式が途中で終わっているため、条件式が解析できません。

(P) 検索条件式を修正して再実行してください。

Incomplete text search condition.

検索条件式が途中で終わっており、検索条件式が解析できません。

(P) 検索条件式を修正して再実行してください。

Incomplete text search condition. offset=aa

検索条件式が途中で終わっています。

aa：検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Index data for search exceeds maximum size.(XX)

参照するインデクスサイズが上限値を超えました。

XX：インデクスサイズの上限値

(P) 検索条件を変更して再実行してください。

Invalid character code.

CREATE TABLE のプラグインオプションでの、文字コード種別の指定に誤りがあります。

(P) プラグインオプションを修正して再実行してください。

Invalid data found.filename= aa..aa,offset=XX,file= bb..bb,line=YY

指定されたインデクスデリミタファイル、不要語ファイル、または削除コードファイル中に不正なデータがあります。または、指定できる個数の上限を超えています。

aa..aa：ファイル名

XX：不正データのデータ先頭からのオフセット

bb..bb：エラーが発生したファイル

YY：エラーが発生したソースファイル行

[対策] データを確認してください。

Invalid delimiter exist. offset=aa

区切り文字が誤っています。

aa：検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Invalid document.

Preprocessing Library for Text Search で作成していない文書を ,utf-8 の HiRDB 運用環境の SGMLTEXT 型の表に登録しようとしてしました。

(P) utf-8 の HiRDB 運用環境で SGMLTEXT 型の表に登録する場合は , Preprocessing Library for Text Search で作成した文書を使用してください。

Invalid extract parameter.line=XX

extracts 関数の指定が誤っています。

XX : 保守用情報

(P) extracts 関数の指定を見直してください。

Invalid extracting structure name. element name=aa..aa.

抽出構造名の指定に誤りがあります。

aa..aa : 要素名

(P) 抽出構造名の指定内容を確認してください。内容を確認後 , 再実行してください。

Invalid Number. offset=aa

有効でない数値が指定されています。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Invalid option. offset=aa

不正なオプションです。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Invalid plugin description.

SGMLTEXT 型の列のプラグインオプションの内容に誤りがあります。

(P) プラグインオプションの内容を確認してください。内容を確認後 , 再実行してください。

Invalid score option name. offset=aa

スコアオプション名の指定に誤りがあります。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Invalid structure name.

構造名の指定に誤りがあります。または , 構造名指定長が 1,024 バイトを超えています。

(P) 検索条件式を修正して再実行してください。

Invalid text data type : aa..aa

テキストデータ種別 (SGMLTEXT 型の列のプラグインオプションのタグ <TEXTTYPE> の内容) の指定に, SGML, PLAIN 以外が指定されました。

aa..aa : 指定された内容

(P) テキストデータ種別を正しく指定して, 再実行してください。

Invalid text search condition.

全文検索条件に誤りがあります。

(P) 検索条件を修正して再実行してください。

Invalid value found in Plugin Option(aa).bb(XX)

プラグインオプションの指定に誤りがあります。

aa : 誤りを発見したパラメタ

bb : エラーの発生したファイル名

XX : エラーの発生した行

[対策] プラグインオプションの指定を修正し, 再度実行してください。

(L)

Left paren symbol not exist after NOT symbol.

NOT 演算子のあとに左括弧がありません。

(P) 検索条件を修正して再実行してください。

(N)

No attribute name. offset=aa

属性名の指定がありません。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

No back slash before escape character. offset=aa

"¥" (エスケープ文字) がありません。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

No delimiter. offset=aa

区切り文字がありません。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

No dictionary name. offset=aa

同義語辞書名の指定がありません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

Not enough shared memory. file=aa..aa,line=XX

aa..aa : オペレーションを発行したソースファイル
XX : オペレーションを発行した行
(P) HiRDB 管理者に連絡してください。
[対策] HiRDB のバックエンドサーバ定義またはシングルサーバ定義の pdplgprm オペランド指定値を見直してください。

No expands mode. offset=aa

異表記展開モードの指定がありません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

No highlight tag in extract parameter.

extracts パラメタにハイライトタグの指定がありません。
(P) HiRDB 管理者に連絡してください。
[対策] extracts パラメタの内容を見直してください。

No number. offset=aa

数値指定がありません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

No structure name. offset=aa

構造名指定がありません。
aa : 検索条件式のオフセット
(P) 検索条件式を修正して再実行してください。

No structure weight after left paren symbol.

構造名指定のあとの、左括弧の後ろの構造重みの指定がありません。
(P) 構造重みを指定して、再実行してください。

No term weight after left paren symbol.

単純文字列指定のあとの、左括弧の後ろにタームの重み指定がありません。
(P) タームの重みを指定して、再実行してください。

No top structure tag in extract parameter.

extracts パラメタの最上位タグの指定がありません。
 (P) HiRDB 管理者に連絡してください。
 [対策] extracts パラメタの内容を見直してください。

Normalize parameter(XX) not catalogued in system.

CREATE TABLE のプラグインオプションに指定した正規化パラメタがレジストりに登録されていません。
 XX: 正規化パラメタ名
 (P) HiRDB 管理者に連絡してください。
 [対策] phssgmlreg (sgml 定義情報登録ユーティリティ) で必要な正規化パラメタを登録してください。

Not Concept Index.

処理対象のインデクスは概念検索用インデクスではありません。
 (P) インデクスの種別を確認してください。

Not enough area.filename=Delay File

分割遅延登録用ファイル領域不足によって、登録に失敗しました。
 (P) HiRDB 管理者に連絡してください。
 [対策] インデクス作成開始ユーティリティで遅延登録文書をすべてインデクスに登録してください。

Not enough memory.file= aa..aa,line= XX

処理に必要なメモリをシステムが取得できませんでした。
 aa..aa: メモリを要求したソースファイル名
 XX: エラーが発生した行
 (P) システムのメモリ状態を確認してください。

Not enough memory for expand search.max memory size= XX

同義語異表記展開処理に必要なメモリが環境情報定義に指定した同義語異表記展開用最大メモリサイズを超えました。
 XX: 環境情報定義に指定した同義語異表記展開用最大メモリサイズ (キロバイト)
 (P) 検索タムを変更するか、または同義語異表記展開の指定を変更して再実行してください。
 [対策] このエラーメッセージが頻繁に出力される場合は、次の対処をして HiRDB を再度起動してください。
 ・ リソースを見積もり直して、環境情報定義の同義語異表記展開用最大メモリサイズの指定値を増やしてください。

Not enough memory for search. max memory size = XX

検索処理に必要なメモリがシステムの上限、または環境情報定義に指定した検索時最大メモリサイズを超えました。

XX：システム上限値または環境情報定義に指定した検索時最大メモリサイズ（キロバイト）

（P）論理演算の条件を変更するか、または異表記展開の指定を変更して再実行してください。

[対策] 環境定義ファイルを使用時に、このエラーメッセージが頻繁に出力されるときは、次の対処をして HiRDB を再度起動してください。

- メモリに余裕があるときは、環境情報定義の検索時最大メモリサイズの指定値を増やしてください。
- メモリに余裕がないときは、環境情報定義の検索オプションに "file" を指定してください。

Not enough RD area. filename=aa..aa

RD エリア容量が不足しています。

aa..aa：容量不足が発生した物理格納定義の種別

（P）HiRDB 管理者に連絡してください。

[対策] RD エリア容量を増やしてください。

Not enough stack memory for search.max memory size= XX

検索オプションに " file " を指定している場合の検索処理に必要なスタックメモリがシステムの上限を超えました。

XX：システム上限値（キロバイト）

（P）頻出語や 1 文字のワイルドカード検索または近傍検索を行わないように条件を変更して再実行してください。

[対策] このエラーメッセージが頻繁に出力されるときは、次の対処をして HiRDB を再度起動してください。

環境情報定義の検索オプションに " memory " を指定してください。

Not enough rdarea size for SGML structure index.

SGMLTEXT 型の列を格納する RD エリアのページ数が少ないため、SGML 構造インデックスが作成できません。

（P）RD エリアのページ数を増やして、再実行してください。または、別の RD エリアに割り当ててください。

Not support language.

HiRDB で設定している文字コードをサポートしていないため処理できません。

（P）Text Search Plug-in がサポートしている文字コードで HiRDB 環境を再構築後、再実行してください。

NOT symbol unable to use at logical term condition.

ターム検索条件の論理演算 (AND/OR) に対して NOT 演算子を指定しています。
 (P) 検索条件を修正して再実行してください。

Number of file is over the system limit.

インデクスファイル数がシステムの上限を超えました。
 (P) HiRDB 管理者に連絡してください。
 [対策] ファイル数がシステムの上限を超えないように、インデクスを作成してください。

Number overflow. offset=aa

数値指定がオーバーフローしました。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

(O)

Overflow of structure weight.

構造の重みの値がオーバーフローしています。
 (P) 1 ~ 100 の値を指定して再実行してください。

Overflow of term weight.

タームの重みの値がオーバーフローしています。
 (P) 1 ~ 100 の値を指定して再実行してください。

(P)

Paren symbol not exist after text function invocation.

近傍条件指定 (PROXIMITY), 展開指定 (SOUNDEX_EXP, SYNONYM) のあとに括弧がありません。
 (P) 検索条件式を修正して再実行してください。

Paren symbol not match in condition.

検索条件式中の括弧が対応していません。
 (P) 検索条件式を修正して再実行してください。

Parenthesis nest overflow. offset=aa

括弧のネストが制限値を超えました。
 aa : 検索条件式のオフセット
 (P) 検索条件式を修正して再実行してください。

Plugin description is not specified at SGMLTEXT type column.

SGMLTEXT 型カラムにプラグインオプションの指定がありません。
(P) 定義内容を見直して再実行してください。

PLUGIN OPTION(aa..aa) Not Installed.

プラグインオプションがインストールされていません。
aa..aa : プラグインオプション名
Concept : Text Search Plug-in Conceptual Extension
[対策] 表示されたプラグインオプションをインストールしてください。

PLUGIN OPTION(CONCEPT_ON) Not defined.

概念検索用のインデクスが定義されていません。
[対策] 検索条件式を修正して再実行してください。

PPI Error occurred.func=aa..aa,code=XX,file=bb..bb,line=YY

HiRDB のサービスコールでエラーが発生しました。
aa..aa : HiRDB サービスコール関数名
XX : 返却されたエラーコード
bb..bb : オペレーションを発行したソースファイル名
YY : オペレーションを発行した行
[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

(R)

Right paren symbol not exist after structure weight.

構造重み指定の後ろに右括弧がありません。
(P) 検索条件式を修正して再実行してください。

Right paren symbol not exist after term weight.

タームの重み指定の後ろに右括弧がありません。
(P) 検索条件式を修正して再実行してください。

(S)

Structure name not found in SGML structure index.

全文検索条件式中に指定された構造名は、SGML 構造インデクスに登録されていません。
(P) 指定した構造名が正しいか確認してください。正しくない場合は、検索条件式を修正して再実行してください。

Synonym dictionary data get failed.

同義語辞書データの取得に失敗しました。

(P) 検索条件中に指定した同義語辞書名が正しいか確認してください。正しくない場合は修正して再実行してください。正しい場合は、HiRDB 管理者に連絡してください。

[対策] 同義語辞書がシステムに登録されているか確認してください。登録されていない場合は、「phssyndicreg (同義語辞書登録ユティリティ)」で同義語辞書に登録してください。

登録されている場合は、メッセージログの内容を確認して、保守員に連絡してください。

Synonym dictionary data not found in system.

指定した同義語辞書は登録されていません。

(P) 指定した同義語辞書名が正しいか確認してください。同義語辞書名が正しい場合は、HiRDB 管理者に連絡してください。

[対策] 同義語辞書に登録してください。

System error. offset=aa condition=bb file=cc line=dd

内部矛盾が発生しました。

aa : 検索条件式のオフセット

bb : 検索条件式

cc : ファイル名

dd : 行番号

[対策] 保守員に連絡してください。

(T)

Term condition not exists.

英文検索でインデクスデリミタ、不要語、削除コードだけで構成されたタームが存在します。

(P) 検索条件式を指定して再実行してください。

Term condition not exist after structure weight.

構造重み指定の後ろにターム条件が指定されていません。

(P) 検索条件式を指定して再実行してください。

Term condition not exists.

ターム条件が指定されていません。

(P) 検索条件式を指定して再実行してください。

Term delimiter(") not exist. offset=aa

" (ダブルクォーテーション) がありません。

aa : 検索条件式のオフセット

(P) 検索条件式を指定して再実行してください。

Text search condition is not specified.

抽象データ型関数 `contains()` , `contains_with_score()` の構造検索条件の指定がありません。

(P) 検索条件式を指定して再実行してください。

The length of concept condition is over limit.

概念検索条件式の長さが制限値を超えました。

(P) 概念検索条件式の長さを制限値以下にしてください。

Token size is over max size(255bytes)

次に示すサイズが指定できる最大数 (255 バイト) を超えています。

単純文字列指定 : " (ダブルクォーテーション) で指定した文字列

同義語辞書名 : SYNONYM で指定している同義語辞書名

構造名 : 構造名指定に指定している構造名

(P) 検索条件式を修正して再実行してください。

Too much term length was specified. offset=aa

指定した長さ (バイト) が正しくありません。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

Too short search factor in specific and. offset=aa

特定構造検索で検索条件が一つしかありません。

aa : 検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

(U)

Unable to execute aa..aa

インデクスに対して、実行できない機能が指定されました。

aa : 機能名

[対策] インデクスの種別を確認してください。

Unable to execute because of RD area rollback

対象となる RD エリアが回復処理中の為、処理できません。

[対策] 対象 RD エリアの回復処理完了後、再実行してください。

Unable to use directory for search work,name=aa bb(XX)

コンフィグレーションファイルに指定された検索用ワークファイルディレクトリが使用できません。

ディレクトリの指定を見直してください。

aa：ディレクトリ名

bb：エラーの発生したファイル名

XX：エラーの発生した行

[対策] コンフィグレーションファイルの指定を見直し、HiRDB を再度起動してください。

Unmatch character code.

表形式・データ登録の場合

utf-8 以外の HiRDB 運用環境で、CREATE TABLE のプラグインオプションに utf-8 用の指定があります。または、表の文字コード種別と登録文書の文字コード種別に整合性がありません。

[対策] 文字コード種別を操作対象に合わせて、指定内容または登録文書を修正して再実行してください。

検索の場合

CREATE TABLE 実行時の文字コード (<Unicode>) 指定値と HiRDB クライアント環境変数 PDCLTCNVMODE の値に整合性がありません。

[対策] CREATE TABLE 実行時の文字コード指定値に合わせて、次のいずれかの値を PDCLTCNVMODE に設定してください。

文字コード指定値が「MS」の場合："UTF8MS" または "NOUSE"

文字コード指定値が「JIS」の場合："UTF8" または "NOUSE"

(W)

Weight exist in specific struct. offset=aa

特定構造検索指定中に、重み指定があります。

aa：検索条件式のオフセット

(P) 検索条件式を修正して再実行してください。

(2) メッセージログに出力されるメッセージ

00001-E PPI Error occurred.func=aa..aa,code=XX,file=bb..bb,line=YY

HiRDB のサービスコールでエラーが発生しました。

aa..aa：HiRDB サービスコール関数名

XX：返却されたエラーコード

bb..bb：オペレーションを発行したソースファイル名

YY：オペレーションを発行した行

[対策] このメッセージの直前にメッセージが出力されている場合は、それによって対処

してください。出力されていない場合は、保守員に連絡してください。

00005-E Not enough memory.file=aa..aa,line=XX

処理に必要なメモリをシステムが取得できませんでした。

aa..aa：メモリを要求したソースファイル名

XX：エラーが発生した行

[対策] システムのメモリ状態を確認してください。

00006-E File open error.filename=aa..aa,code=XX

aa..aa に示されるファイルのオープンに失敗しました。

aa..aa：オープンできなかったファイル名

XX：詳細情報

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

00007-E File I/O error.filename=aa..aa,func=bb..bb,code=XX,file=cc..cc,line=YY

システムファイルのアクセス中にエラーを検出しました。

aa..aa：ファイル名

bb..bb：オペレーション

XX：詳細情報

cc..cc：オペレーションを発行したソースファイル名

YY：オペレーションを発行した行

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

00008-E Fatal error occurred.detail=aa..aa,file=bb..bb,line=XX

システム内部で自己矛盾を検出しました。

aa..aa：詳細情報

bb..bb：エラーの発生したファイル名

XX：エラーの発生した行

[対策] このメッセージの直前にメッセージが出力されている場合は、それに従って対処してください。出力されていない場合は、保守員に連絡してください。

00009-E Not enough memory for search.max memory size= XX

検索処理に必要なメモリがシステムの上限または環境情報定義に指定した検索時最大メモリサイズを超えました。

XX：システム上限値または環境情報定義に指定した検索時最大メモリサイズ(キロバイト)

(P) 論理演算の条件を変更するか、または異表記展開の指定を変更して再実行してください。

[対策] 環境定義ファイルを使用時に、このエラーメッセージが頻繁に出力されるとき

は、次の対処をして HiRDB を再度起動してください。

- メモリに余裕があるときは、環境情報定義の検索時最大メモリサイズの指定値を増やしてください。
- メモリに余裕がないときは、環境情報定義の検索オプションに "file" を指定してください。

00010-E Not enough RD area.filename=aa..aa

RD エリア容量が不足しています。

aa..aa : 容量不足が発生した物理格納定義の種類別

[対策] RD エリア容量を増やしてください。RD エリアは、HiRDB の「pdmod (データベース構成変更ユティリティ)」を実行して追加します。また、インデクス領域として使用する場合は、「phnexpnd (インデクスの拡張ユティリティ)」または「phxexpnd (インデクスの拡張ユティリティ)」を実行してください。

00011-E File size is over the system limit. filename = aa..aa

aa..aa に示されるファイル容量がシステムの上限を超えました。

aa..aa : サイズがシステムの上限を超えたファイル名

(P) HiRDB 管理者に連絡してください。

[対策] ファイルサイズがシステムの上限を超えないように、インデクスを作成してください。

00012-E Config File I/O

error.filename=aa..aa,func=bb..bb,code=XX,file=cc..cc,line=YY

コンフィグレーションファイルのアクセス中にエラーを検出しました。

aa..aa : ファイル名

bb..bb : オペレーション

XX : 詳細情報

cc..cc : オペレーションを発行したソースファイル名

YY : オペレーションを発行した行

[対策] このメッセージの直前にメッセージが出力されている場合は、それに従って対処してください。出力されていない場合は、保守員に連絡してください。

00013-E Number of file is over the system limit.

インデクスファイル数がシステムの上限を超えました。

[対策] ファイル数がシステムの上限を超えないインデクスを作成してください。

00014-E Error in Plugin Option aa(XX)

プラグインオプションの指定に誤りがあります。

aa : エラーの発生したファイル名

XX : エラーの発生した行

[対策] プラグインオプションの指定を修正し、再度実行してください。

00015-E Invalid value found in Plugin Option(aa).bb(XX)

プラグインオプションの指定に誤りがあります。または、utf-8 の HiRDB 運用環境で使用できないオプションを指定しています。

aa : 誤りを発見したパラメタ

bb : エラーの発生したファイル名

XX : エラーの発生した行

[対策] プラグインオプションの指定を修正し、再度実行してください。

00016-E Unable to use directory for search work,name=aa bb(XX)

コンフィグレーションファイルに指定された検索用ワークファイルディレクトリが使用できません。

ディレクトリの指定を見直してください。

aa : ディレクトリ名

bb : エラーの発生したファイル名

XX : エラーの発生した行

[対策] コンフィグレーションファイルの指定を見直し、HiRDB を再度起動してください。

00017-E PLUGIN OPTION(aa..aa) Not Installed.

プラグインオプションがインストールされていません。

aa..aa : プラグインオプション名

Concept : Text Search Plug-in Conceptual Extension

[対策] 表示されたプラグインオプションをインストールしてください。

00018-E Define file not exist.filename= aa..aa,func=bb..bb,file=cc..cc,line=XX

指定されたインデクスデリミタファイル、不要語ファイル、または削除コードファイルが存在しません。

aa..aa : ファイル名

bb..bb : オペレーション

cc..cc : オペレーションを発行したソースファイル

XX : オペレーションを発行したソースファイル行

[対策] プラグインオプションに指定したファイル名称を確認してください。

00019-E Invalid data found.filename= aa..aa,offset=XX,file= bb..bb,line=YY

指定されたインデクスデリミタファイル、不要語ファイル、または削除コードファイル中に不正なデータが存在します。

aa..aa : ファイル名

XX : 不正データのデータ先頭からのオフセット

bb..bb : エラーが発生したソースファイル

YY : エラーが発生したソースファイル行

[対策] データを確認してください。

00020-E Fail to lock aa..aa.

aa..aa の排他に失敗しました。n-gram インデクス用 RD エリアまたは SGMLTEXT データ用 RD エリアへのアクセスが競合しています。

aa..aa : Index File (n-gram インデクス用 RD エリア)

SGML Logical File (SGMLTEXT データ用 RD エリア)

[対策] しばらく待って再実行してください。このメッセージの直前または直後にメッセージが出力されている場合は、そのメッセージの内容に従って対処してください。

00021-E Unable to execute because of RD area rollback.

対象となる R D エリアが回復処理中の為、処理できません。

[対策] 対象 R D エリアの回復処理完了後、再実行してください。

00024-E Not enough shared memory. file= aa..aa, line=bb..bb.

プラグインに割り当てられた共用メモリが不足しています。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行

[対策] HiRDB のバックエンドサーバ定義またはシングルサーバ定義の「pdplgrm」オペランド指定値を見直してください。

00026-E Fail to lock because file is over the process limit. filename=aa..aa, line=bb..bb.

1 プロセスのオープンファイル数が上限を超えました。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行

[対策] しばらく待って再実行してください。再実行の際にも同様のエラーが発生した場合は、プロセスごとに対応してください。

- HiRDB のサーバプロセスの場合
HiRDB 管理者に連絡してください。
- ユーザ起動によるプロセス (コマンド・ユティリティ) の場合
OS のパラメタの maxfiles の設定値を変更してください。

00027-E Fail to lock because file is over the system limit. filename=aa..aa, line=bb..bb.

システム全体のオープンファイル数が上限を超えました。

aa..aa : ファイル名

bb..bb : オペレーションを発行した行

[対策] しばらく待って再実行してください。再実行の際にも同様のエラーが発生した場合は、次に示すように対応してください。

- OS のパラメタの nfile の設定値を変更する。
- コマンドを実行したサーバマシンで動作している不要なプロセスを停止する。
- コマンドを実行しない。

00028-E Fail to thread lock. filename=aa..aa, line=bb..bb.

排他に失敗しました。排他エリアへのアクセスが競合しています。
aa..aa：ファイル名
bb..bb：オペレーションを発行した行
[対策]しばらく待って再実行してください。

00029-E Shared memory is not defined in HiRDB definition. file=aa..aa, line=bb..bb.

プラグイン用の共用メモリが HiRDB サーバ定義に指定されていません。
aa..aa：ファイル名
bb..bb：オペレーションを発行した行
[対策]プラグイン用の共用メモリを HiRDB サーバ定義に指定してください。

00030-E Not enough area.filename=Delay File

分割遅延登録用ファイル領域不足によって、登録に失敗しました。
(P) HiRDB 管理者に連絡してください。
[対策]インデクス作成開始ユティリティで遅延登録文書をすべてインデクスに登録してください。

00031-E Not enough stack memory for search.max memory size= XX

検索オプションに "file" を指定している場合の検索処理に必要なスタックメモリがシステムの上限を超えました。
XX：システム上限値 (キロバイト)
(P) 頻出語や 1 文字のワイルドカード検索または近傍検索を行わないように条件を変更して再実行してください。
[対策]このエラーメッセージが頻繁に出力されるときは、次の対処をして HiRDB を再度起動してください。
環境情報定義の検索オプションに "memory" を指定してください。

00032-E Not Concept Index.

処理対象のインデクスは概念検索用インデクスではありません。
[対策]インデクスの種別を確認してください。

00033-E Directory or File for Concept Information not exist or destroyed.

概念情報として指定したディレクトリもしくはファイルが存在しないか、または破壊されています。
[対策]指定したディレクトリまたはファイルを確認してください。

00034-E Invalid Index Type.

指定したインデクス種別はサポートしていません。
(P) インデクスの種別を確認してください。

00035-E Index data for search exceeds maximum size.(XX)

上限値を超えるインデクスを使用した検索が実行されました。

XX：インデクスサイズの上限值

[対策] ひらがなやかたかな 1 文字などの、システムに高負荷を与える検索タームを指定しています。検索タームを見直して再実行してください。

00037-E Not enough memory for expand search.max memory size= XX

同義語異表記展開処理に必要なメモリが環境情報定義に指定した同義語異表記展開用最大メモリサイズを超えました。

XX：環境情報定義に指定した同義語異表記展開用最大メモリサイズ（キロバイト）

(P) 検索タームを変更するか、または同義語異表記展開の指定を変更して再実行してください。

[対策] このエラーメッセージが頻繁に出力されるときは、次の対処をして HiRDB を再度起動してください。

- リソースを見積もり直して、環境情報定義の同義語異表記展開用最大メモリサイズの指定値を増やしてください。

00038-E Each definition unmatched. value1=Index Option aa, value2=bb

aa と bb の定義の組み合わせは、実行できない組み合わせです。

aa：CREATE INDEX のプラグインオプションに指定したインデクス作成方法

bb：Lockmode または -t time 指定

[対策] aa と bb の定義内容を見直してください。

| aa | bb | 対策 |
|--------------|----------|--|
| INDEX_UCS4 | -t time | UCS4 インデクスには -t 指定はできません |
| CREATEMODE=3 | Lockmode | データ型プラグインの環境定義ファイルの phs_search_wait または phc_search_wait に「nowait_with_update」が指定されているかどうかを確認してください。また、インデクス作成開始ユティリティ (phnstartidx または phxstartidx) の -w オプションに、「nowait_with_update」が指定されているかどうかを確認してください。 |

00040-E Unable to use function. function name=INDEX_UCS4.

サポート外機能です。

[対策] phn(x)modidx コマンドで UCS4 インデクスへ移行する際に、UCS4 インデクスがサポートしていない機能を使用したインデクスではないかを確認してください。

01001-W Invalid data found.file=aa,line=XX.bb(YY)

誤りのある行がコンフィグレーションファイルにあります。誤りを発見したコンフィグレーションファイルの指定内容をすべて無効とします。

aa：誤りを発見したコンフィグレーションファイル名

XX：誤りを発見したコンフィグレーションファイルの行番号

bb : エラーの発生したファイル名

YY : エラーの発生した行

[対策] コンフィグレーションファイルの内容を修正し、HiRDB を再度起動してください。

01002-W Invalid value found. file=aa,line=XX,item=bb.cc(YY)

コンフィグレーションファイルの指定内容に誤りがあります。誤りを発見したコンフィグレーションファイルの指定内容をすべて無効とします。

aa : 誤りを発見したコンフィグレーションファイル名

XX : 誤りを発見したコンフィグレーションファイルの行番号

bb : 誤りを発見した指定項目

cc : エラーの発生したファイル名

YY : エラーの発生した行

[対策] コンフィグレーションファイルの内容を修正し、HiRDB を再度起動してください。

01003-W Not enough RD area for expand.

インデクスを拡張するための RD エリアが不足しています。インデクスの拡張処理を実行しません。

[対策] RD エリアを拡張後、再実行してください。

01004-W Unable to execute aa..aa

インデクスに対して実行できない機能が指定されました。

aa : 機能名

[対策] インデクスの種別を確認してください。

01005-W Specified size not enough for expand.

指定されたサイズが小さいため、ファイルが拡張できません。

[対策] 拡張するファイルが分割遅延登録用ファイルである場合、拡張できる最小単位は 2,048 キロバイトです。そのため、現在の分割遅延登録用ファイルサイズ + 2,048 キロバイト以上を指定してください。

02000-I Sub index using ratio XX percent

差分インデクスの使用率が XX パーセントになりました。

XX : 差分インデクスの使用率

02001-I aa..aa is Time Out.

コマンドが指定時間を過ぎて終了しました。

aa..aa : コマンド名

02002-I aa..aa command started.

コマンドを開始しました。
aa..aa : コマンド名

02003-I aa..aa command ended.

コマンドを終了しました。
aa..aa : コマンド名

02004-I Delay file using ratio XX percent.

分割遅延登録用ファイル使用率が XX パーセントになりました。
XX : 分割遅延登録用ファイル使用率

02005-I To have processed in the full counts, aa has finished.

コマンドに指定された件数の登録が完了しました。
aa..aa : コマンド名

02006-I XX Processed.

XX 件のデータをインデクス作成しました。
XX : インデクスを作成したデータ件数

02007-I Unable to make index complete. Because of Lock.

作成要求された件数に達する前に、処理できる件数が終わりました。

50001-E Fatal error occurred.

回復不能エラーが発生しました。
[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

50002-E Not enough memory.size=XX file=aa..aa line=YY

メモリ不足が発生しました。
XX : 要求したサイズ
aa..aa : エラーが発生したファイル名
YY : エラーが発生した場所
[対策] システムのメモリ状態を確認してください。

50003-E Registry data get failed. Key name=aa..aa code=XX

レジストリ登録情報の取得でエラーが発生しました。
aa..aa : レジストリ登録情報名
XX : エラーコード
[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。また、レジストリ機能の初期設定を実施しているか確認してください。

何もメッセージが出力されていない場合は保守員に連絡してください。

50004-E Initialize SGML structure index failed.code=XX file=aa..aa line=YY

SGML 構造インデクスの初期化に失敗しました。

XX：エラーコード

aa..aa：エラーの発生したファイル名

YY：エラーの発生した場所

[対策] このメッセージの直前に、別のメッセージが出力されている場合はそのメッセージに従って対処してください。それ以外は、保守員に連絡してください。

50005-E SGML compiler error occurred.file=aa..aa line=XX

SGML 文書の解析中にエラーが発生しました。

aa..aa：エラーの発生したファイル名

XX：エラーの発生した場所

[対策] 保守員に連絡してください。

50006-E SGML structure index open failed.code=XX file=aa..aa line=YY

SGML 構造インデクスのオープンでエラーが発生しました。

XX：エラーコード

aa..aa：エラーが発生したファイル名

YY：エラーが発生した場所

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。何もメッセージが出力されていない場合は、保守員に連絡してください。

50007-E SGML structure index add failed.code=XX file=aa..aa line=YY

SGML 構造インデクスの追加でエラーが発生しました。

XX：エラーコード

aa..aa：エラーが発生したファイル名

YY：エラーが発生した場所

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。何もメッセージが出力されていない場合は、保守員に連絡してください。

50008-E Plain text compiler error occurred.file=aa..aa line=XX

プレーンテキストの解析中に Text Search Plug-in でエラーが発生しました。

aa..aa：エラーの発生したファイル名

XX：エラーの発生した場所

[対策] 保守員に連絡してください。

50009-E SGML Compiler Message file open failed.

Text Search Plug-in が必要なメッセージファイルのオープンに失敗しました。

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処し

てください。何もメッセージが出力されていない場合は、Text Search Plug-in が正しく登録されているか確認してください。

50010-E SPK Message file open failed.

Text Search Plug-in が必要なメッセージファイルのオープンに失敗しました。

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。何もメッセージが出力されていない場合は、Text Search Plug-in が正しく登録されているか確認してください。

50011-E PPI Error occurred. func=aa...aa code=XX,file=bb..bb,line=YY

HiRDB のサービスコールでエラーが発生しました。

aa..aa : HiRDB サービスコール関数名

XX : 返却されたエラーコード

bb..bb : エラーの発生したファイル名

YY : エラーの発生した場所

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

50013-E Katakana rule data get failed.

カタカナルールファイルの読み込みに失敗しました。

[対策] このメッセージの直前に、別のメッセージが出力されている場合はそのメッセージに従って対処してください。それ以外は、保守員に連絡してください。

50014-E Logical file I/O error occurred.func=aa..aa code=XX

論理ファイル操作でエラーが発生しました。

[対策] このメッセージの直前に、別のメッセージが出力されている場合はそのメッセージに従って対処してください。それ以外は、保守員に連絡してください。

50015-E Synonym dictionary data get failed.func=aa..aa code=XX file=bb..bb line=YY

同義語辞書の読み込みに失敗しました。

aa..aa : 関数名

XX : エラーコード

bb..bb : エラーの発生したファイル名

YY : エラーの発生した場所

[対策] このメッセージの直前に、別のメッセージが出力されている場合はそのメッセージに従って対処してください。それ以外は、保守員に連絡してください。

50016-E Invalid extract parameter. line=XX

extracts パラメタの内容が誤っています。

XX : 誤りのある行

[対策] extracts パラメタの内容を見直してください。

50017-E No highlight tag in extract parameter.

extracts パラメタにハイライトタグの指定がありません。

[対策] extracts パラメタの内容を見直してください。

50018-E No top structure tag in extract parameter.

extracts パラメタに最上位タグの指定がありません。

[対策] extracts パラメタの内容を見直してください。

50019-E DTD name for display is not specified in extract parameter.

extracts パラメタに表示用 DTD 名の指定がありません。

[対策] extracts パラメタの内容を見直してください。

50020-E Specified character code set not supported.

指定されている文字コード種別はサポートしていません。

[対策] HiRDB システムの OS への登録時に指定した文字コード種別を見直してください。

62043-E Specified character code set not supported.

未サポート文字コードの HiRDB 運用環境に Text Search Plug-in をセットアップして HiRDB を起動しました。

[対策] Text Search Plug-in がサポートしている文字コードの HiRDB 運用環境でセットアップしてください。

(3) 標準エラー出力に出力されるメッセージ

52001-I aa..aa command started.

aa..aa に示すコマンドを開始しました。

aa..aa : コマンド名

[対策] なし

52002-I aa..aa command ended.

aa..aa に示すコマンドが正常終了しました。

aa..aa : コマンド名

[対策] なし

52003-E aa..aa command error occurred.code=XX

aa..aa に示すコマンドが異常終了しました。

aa..aa : コマンド名

XX : エラーコード

[対策] 直前に出力されているメッセージに従って対処してください。

52004-E System call error occurred.func=aa..aa code=XX

システムコールでエラーが発生しました。

aa..aa : システムコール名

XX : エラーコード

[対策] OS のマニュアルに従って対処してください。

52005-E Illegal options - XX

オプションの指定が間違っています。

XX : 指定されたオプションの内容

[対策] 正しいオプションを指定して再実行してください。

52006-I Usage : phsregsetup [-d]

コマンドの形式を示します。

[対策] なし

52007-W Context for DTD already exist in registry.continue..

DTD 用のコンテキストがすでにあります。

phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。

[対策] なし

52008-W Context for normalize parameters already exist in registry.continue..

正規化パラメタ用のコンテキストがすでにあります。

phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。

[対策] なし

52009-W Context for extract parameters already exist in registry.continue..

extracts パラメタ用のコンテキストがすでにあります。

phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。

[対策] なし

52010-W Context for DTD not found in registry.continue..

DTD 用のコンテキストが見つかりません。

phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。

[対策] なし

52011-W Context for extract parameters not found in registry.continue..

extracts パラメタ用のコンテキストが見つかりません。

phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。

[対策] なし

52012-W Context for normalize parameters not found in registry.continue..

正規化パラメタ用のコンテキストが見つかりません。
phsregsetup (レジストリ初期化ユティリティ) は処理を続行します。
[対策] なし

52013-E Communication error occurred.

HiRDB と接続できませんでした。
[対策] 直前に出力されている HiRDB のメッセージに従って対処してください。

52014-I Context for DTD created.

DTD 用のコンテキストを作成しました。
[対策] なし

52015-I Context for normalize parameters created.

正規化パラメタ用のコンテキストを作成しました。
[対策] なし

52016-I Context for extract parameters created.

extracts パラメタ用のコンテキストを作成しました。
[対策] なし

52017-I Context for DTD deleted.

DTD 用コンテキストを削除しました。
[対策] なし

52018-I Context for normalize parameters deleted.

正規化パラメタ用のコンテキストを削除しました。
[対策] なし

52019-I Context for extract parameters deleted.

extracts パラメタ用のコンテキストを削除しました。
[対策] なし

**52020-I Usage : phssgmlreg <AttributeName> <AttrFileName>
<FilePathName>**

コマンドの形式を示します。
[対策] なし

52021-E Illegal attribute name - aa..aa

定義属性名の指定に誤りがあります。
[対策] 正しい定義属性名を指定して再実行してください。

52022-E Not found FilePathName - aa..aa

実体ファイル名に指定したファイルが見つかりません。

[対策] 正しい実体ファイル名を指定して再実行してください。

52023-E AttrFileName is too long.

定義属性名が最大長を超えています。

[対策] 定義ファイル名を 240 バイト未満にして再実行してください。

52024-E Not enough memory.

メモリ不足が発生しました。

[対策] 不要なプログラムを終了するなどしてメモリを確保し、再実行してください。

52025-E Context for SGML plugin not found in registry.enter phsregsetup command.

SGML プラグインで使用するコンテキストがレジストリに登録されていません。

[対策] phsregsetup (レジストリ初期化ユティリティ) でレジストリを初期化したあとに再実行してください。

52026-Q 'aa..aa' is already exist in registry.overwrite?(y/n default n)

指定された定義属性名がすでに登録されているため、上書きしてもよいか問い合わせます。

aa..aa : 定義属性名

[対策] 上書きする場合は、"y" を入力します。処理をキャンセルする場合は、その他を入力します。

52027-I 'aa..aa' is catalogued to registry.

指定された定義属性名がレジストリに登録されました。

aa..aa : 定義属性名

[対策] なし

52028-I Usage : phssgmlrm <AttributeName> <AttrFileName>

コマンドの形式を示します。

[対策] なし

52029-E 'aa..aa' is not catalogued in registry.

指定された定義属性名がレジストリに登録されていません。

aa..aa : 定義属性名

[対策] 正しい定義属性名を指定して再実行してください。

52030-I 'aa..aa' is deleted from registry.

指定された定義属性名がレジストリから削除されました。

aa..aa : 定義属性名

[対策] なし

52031-I Usage:phssgmlval<AttributeName><AttrFileName>[<Output File>]

コマンドの形式を示します。

[対策] なし

52032-E Attribute file data exceeds maximum size.

SGML 定義情報のデータサイズが制限値 (5 メガバイト) を超えています。

[対策] SGML 定義情報のデータサイズを制限値 (5 メガバイト) 以下にして再実行してください。

52033-E AttrFileName length is 0 byte.

指定した定義ファイル名の長さが 0 バイトです。

[対策] 定義ファイル名の長さを有効長 (1 ~ 240 バイト) の範囲で指定し、再実行してください。

52034-E File path name length is 0 byte.

指定したファイル名の長さが 0 バイトです。

[対策] ファイル名を正しく指定して再実行してください。

52035-E Output file path name length is 0 byte.

指定した出力ファイル名の長さが 0 バイトです。

[対策] 出力ファイル名を正しく指定して再実行してください。

52036-E HiRDB library error occurred.func=aa..aa code= XX file=bb..bb line=YY

HiRDB クライアントライブラリ関数でエラーが発生しました。

aa..aa : 関数名

XX : エラーコード

bb..bb : ファイル名

YY : エラーが発生した行

[対策] 直前に出力されている HiRDB のメッセージに従って対処してください。

52301-E Continuous comma exist.line=XX

"," (半角コンマ) が連続して存在しています。

または、同義語辞書群の中に "¥" (エスケープ文字) だけの同義語辞書があります。"¥" を同義語とする場合、"¥¥" と記述してください。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52302-E Text data ended with comma.line=XX

同義語辞書テキストファイルが","(半角コンマ)で終わっています。

または、同義語辞書群の中に"¥"(エスケープ文字)だけの同義語辞書があります。"¥"を同義語とする場合、"¥¥"と記述してください。

XX: エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52303-E File aa..aa open failed.(code:XX)

ファイルのオープンに失敗しました。

aa..aa: ファイル名

XX: エラーコード

[対策] OSのマニュアルを参照の上、対処してください。

52304-E File aa..aa read failed.(code:XX)

ファイルからの読み込みに失敗しました。

aa..aa: ファイル名

XX: エラーコード

[対策] OSのマニュアルを参照の上、対処してください。

52305-E File aa..aa remove failed.(code:XX)

ファイルの削除に失敗しました。

aa..aa: ファイル名

XX: エラーコード

[対策] OSのマニュアルを参照の上、対処してください。

52306-E File aa..aa write failed.(code:XX)

ファイルの書き込みに失敗しました。

aa..aa: ファイル名

XX: エラーコード

[対策] OSのマニュアルを参照の上、対処してください。

52307-E Illegal characters exist.line=XX

同義語辞書テキストファイルに誤った文字があります。

XX: エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52308-E Invalid file path name (aa..aa).

同義語辞書テキストファイルのパス名が誤っています。

aa..aa: 同義語辞書テキストファイルのパス名

[対策] 正しいパス名を指定して再実行してください。

52309-E Synonym dictionary file aa..aa not found in system.

同義語辞書はシステムに登録されていません。

aa..aa : 同義語辞書名

[対策] システムに登録されている同義語辞書名を指定して再実行してください。

52310-E Synonym data not found.

同義語辞書テキストファイルに登録が必要なデータが存在しません。

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52311-E Not enough memory.file=aa..aa line=XX

メモリ不足が発生しました。

aa..aa : エラーの発生したファイル名

XX : エラーの発生した場所

[対策] しばらく待って再実行してください。

52312-E Not SJIS code character exist.line=XX

同義語辞書テキストファイル中に半角文字があります。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52313-E Only one word in synonym group.line=XX

同義語群の中に同義語が一つしか指定されていません。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52314-E Num of character's kind over 1000.line=XX

文字種が 1,000 を超えました。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52315-E Text data started with comma.line=XX

同義語辞書の先頭カラムに "," (半角コンマ) が存在する行があります。または、同義語辞書群の中に "¥" (エスケープ文字) だけの同義語があります。"¥" を同義語とする場合、"¥¥" と記述してください。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52316-I Usage : aa..aa <dictionary name> <file path name>

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phssyndicreg または phcsyndicreg)

[対策] なし

52317-I Usage : aa..aa <dictionary name>

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phssyndicrm または phcsyndicrm)

[対策] なし

52318-Q aa..aa is already exist.overwrite?(y/n default n)

同義語辞書はすでに登録されているので、上書きしてもよいか問い合わせます。

aa..aa : 同義語辞書名

[対策] 上書きして登録する場合は "y", 処理を中断する場合はその他を入力してください。

52319-E Synonym dictionary name is too long.

同義語辞書名が最大長を超えています。

[対策] 同義語辞書名を 32 バイト以下にして再実行してください。

52320-E Synonym dictionary name length is 0 byte.

指定した同義語辞書名の長さが 0 バイトです。

[対策] 同義語辞書名を有効長 (1 ~ 32 バイト) の範囲で指定し、再実行してください。

52321-E Invalid character is specified in synonym dictionary name.

指定した同義語辞書名に誤った文字が使用されています。

[対策] 同義語辞書名を正しく指定して再実行してください。

52322-E Tab code exist. line=XX

同義語辞書テキストファイル中にタブコードがあります。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

52323-E Illegal character exist after back slash symbol. line=XX

同義語辞書テキストファイル中に、"¥" (エスケープ文字) の後ろに誤った文字が存在する行があります。

XX : エラーの発生した行

[対策] 同義語辞書テキストファイルを修正して再実行してください。

62001-I aa..aa command started.

ユティリティコマンドが開始しました。

aa..aa : ユティリティコマンド名

[対策] なし

62002-I aa..aa command ended.

ユティリティコマンドが正常終了しました。

aa..aa : ユティリティコマンド名

[対策] なし

62003-E aa..aa command error occurred.code=XX

ユティリティコマンドが異常終了しました。

aa..aa : ユティリティコマンド名

XX : エラーコード

[対策] 直前に出力されているメッセージに従って対処してください。

62004-E Environment value(aa..aa) get failed.

環境変数の取得に失敗しました。

aa..aa : 環境変数名

[対策] メッセージに出力された環境変数を設定して再実行してください。

62005-E System call error occurred.func=aa..aa code=XX

システムコールでエラーが発生しました。

aa..aa : システムコール名

XX : エラーコード

[対策] OS のマニュアルを参照の上、対処してください。

62006-E Not enough memory.

メモリ不足が発生しました。

[対策] 不要なプログラムを終了するなどしてメモリを確保し、再実行してください。

62007-E Error occurred in n-gram index plugin.

n-gram インデクスプラグインでエラーが発生しました。

[対策] 直前に出力されているメッセージに従って対処してください。

62008-E HiRDB command (aa..aa) not found in system.

HiRDB のコマンドが見つかりません。

aa..aa : HiRDB コマンド名

[対策] HiRDB が正しくセットアップされているか、またはパスに \$PDDIR/bin が設定されているか確認してください。確認後、HiRDB を再セットアップ、または \$PDDIR/bin をパスに設定して再実行してください。

62009-E HiRDB command (aa..aa) error occurred.code=XX

HiRDB のコマンドでエラーが発生しました。

aa..aa : HiRDB コマンド名

XX : エラーコード

[対策] 直前に出力されているメッセージに従って対処してください。

62010-I Usage : aa..aa -d index_name -r rdarea_name [-l log_mode] [-t time_out_minutes] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnincmt または phxincmt)

[対策] なし

62011-I Usage : aa..aa -d index_name -r rdarea_name [-w wait_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnidxls または phxidxls)

[対策] なし

62012-E Index name length is 0 byte.

指定したインデクス名の長さが 0 バイトです。

[対策] インデクス名を正しく指定し、再実行してください。

62013-E "RDAREA name length is 0 byte.

指定した RD エリア名の長さが 0 バイトです。

[対策] RD エリア名を正しく指定し、再実行してください。

62014-E Option argument invalid. option=xx

オプションに指定された引数 xx が誤っています。

xx : オプション名

[対策] 正しい引数を指定して、再度実行してください。

62015-I Usage : phnmerge -d index_name [-l log_mode] [-w wait_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

[対策] なし

62016-I Usage : aa..aa -d index_name -r rdarea_name [-l log_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnexpnd または phxexpnd)

[対策] なし

62017-I Usage : aa..aa -d index_name [-t time] [-l log_mode] [-c count] [-w wait mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnstartidx または phxstartidx)

[対策] なし

62018-Q Do you really want to make index without time option ? (y/n default n)

実行時間が指定されていません。

[対策] y を入力した場合，時間制限なしで動作します。n を入力した場合，処理を中断します。

62019-Q All Documents of Unfinished Index are Processed, Is it OK? (y/n default n)

分割遅延登録対象文書をすべてインデクスに登録します。

[対策] y を入力した場合，分割遅延登録対象文書をすべてインデクスに登録します。n を入力した場合，処理を中断します。

62020-I Usage : aa..aa -d index_name -r rdarea_name [-l log_mode] [-t time_out_minutes] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phncond または phxcond)

[対策] なし

62021-I Usage : aa..aa -d index_name -v value [-l log_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnmodidx または phxmodidx)

[対策] なし

62022-I Usage : aa..aa -d index_name -m delay_mode [-l log_mode] [-f] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phndelay または phxdelay)

[対策] なし

62023-E Usage : phnchgcon -d index_name { -r rdarea_name | -m termget_option } [-l log_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnchgcon)

[対策] なし

62024-I Usage : aa..aa -d index_name [-r rdarea_name] -o file_name [-w wait_mode] -m collect_mode [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnconget)

[対策] なし

62025-I Usage : aa..aa -i file_name -o file_name

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnconbuild)

[対策] なし

62026-I Usage : aa..aa -d index_name [-r rdarea_name] -i file_name [-l log_mode] [-q generation_num]

ユティリティコマンドの形式を示します。

aa..aa : ユティリティコマンド名 (phnconrstr)

[対策] なし

62027-E File name length is 0 byte.

ファイル名の長さが 0 バイトです。

[対策] 指定したファイル名を見直してください。

62028-E File size error.

ファイルサイズが不正です。

[対策] 指定したファイル名を見直してください。

62029-E File format error.

ファイル内容が不正です。

[対策] 指定したファイルを見直してください。

62030-E Not enough memory.file=aa..aa,line=XX ¥n"

メモリ不足が発生しました。

aa..aa : エラーの発生したファイル名

XX : エラーの発生した場所

[対策] しばらく待って再実行してください。

62031-E File aa..aa open Failed. (code:XX) file=bb..bb line=YY

ファイルのオープンに失敗しました。

aa..aa : ファイル名

XX : エラーコード

bb..bb : エラーの発生したファイル名

YY : エラーの発生した場所

[対策] OS マニュアルを参考の上、対処してください。

62032-E File I/O error.filename=aa..aa,func=bb..bb,code=XX,file=cc..cc,line=YY

システムファイルのアクセス中にエラーを検出しました。

aa..aa : ファイル名

bb..bb : オペレーション

XX：詳細情報

cc..cc：オペレーションを発行したソースファイル名

YY：オペレーションを発行した行

[対策] このメッセージの直前にメッセージが出力されている場合はそれに従って対処してください。出力されていない場合は、保守員に連絡してください。

62033-E Fatal error occurred.detail=aa..aa,file=bb..bb,line=XX

システム内部で自己矛盾を検出しました。

aa..aa：詳細情報

bb..bb：エラーの発生したファイル名

XX：エラーの発生した行

[対策] このメッセージの直前にメッセージが出力されている場合は、それに従って対処してください。出力されていない場合は、保守員に連絡してください。

62034-E Filename or status error. filename=aa..aa

ファイル名またはパス名が不正です。

aa..aa：ファイル名

[対策] 指定したファイル名、またはパス名を見直してください。

62035-E Communication error occurred.

HiRDB と接続できませんでした。

[対策] 直前に出力されている HiRDB のメッセージに従って対処してください。

62036-E HiRDB library error occurred.func=aa..aa code=XX file=bb..bb line=YY

HiRDB クライアントライブラリ関数でエラーが発生しました。

aa..aa：関数名

XX：エラーコード

bb..bb：ファイル名

YY：エラーが発生した行

[対策] 直前に出力されている HiRDB のメッセージに従って対処してください。

62037-E File not found in default directory.

定義ファイルの DEFAULT に指定したパスにファイルが存在しません。

[対策] DEFAULT に指定したパスにファイルが存在するか見直してください。

62038-E File or Directory not found. filename=aa..aa

指定したファイル名、またはパス名が見つかりません。

aa..aa：ファイル名

[対策] 指定したファイル名、またはパス名が存在するか見直してください。

62039-E File or Directory already exist. filename=aa..aa

指定したファイル名, またはパス名はすでに存在します。

aa..aa : ファイル名

[対策] 存在しないファイル名, またはパス名を指定してください。

62040-E Directory or File for Concept Information not exist or destroyed.

指定したパス, またはファイルが存在しないか, 破壊されています。

[対策] 指定されたパス名, またはファイル名を見直してください。

62043-E Specified character code set not supported.

utf-8 の HiRDB 運用環境で使用できないユティリティコマンドを実行しました。

[対策] HiRDB の運用環境を見直してください。

99999-E TSPlugin version unmatched.

SGML プラグイン /n-gram インデクスプラグイン間のバージョンが不整合です。または, FREEWORD プラグイン /IXFREEWORD プラグイン間のバージョンが不整合です。

[対策] Text Search Plug-in を再登録してください。

付録 F.5 SGML 文書の登録エラーメッセージ

SGML 文書の登録時のエラーは, SQL 連絡領域やエラー情報ファイル (pdload 実行時のエラーの場合) に出力されます。各エラーの出力形式について次に説明します。

(1) SQL 連絡領域にメッセージが返却される場合

SQL 連絡領域にメッセージが返却される場合, エラーコードの詳細化指定をしているかどうかによって, 返却されるメッセージに違いがあります。

(a) エラーコードの詳細化指定をしない場合

エラーコードの詳細化指定をしない場合, SQL 連絡領域にメッセージが返却される時は, HiRDB のメッセージ KFPA11997-E とともに出力されます。出力される形式は次のとおりです。なお, エラーコードの詳細化指定をするかどうかは, 環境定義ファイルで指定できます。

KFPA11997-E Error occurred during execution of PLUGIN function, "aa....aa"

aa....aa : (ID=xxxxx) エラー内容

エラー内容が出力されます。

ID=xxxxx の形式で Text Search Plug-in でのメッセージ ID が出力されます。

メッセージ ID の意味については, 「付録 F.6 SGML 文書の登録エラーメッセージ一覧」を参照してください。

KFPA11997-E のメッセージの詳細については、マニュアル「HiRDB メッセージ」を参照してください。

(b) エラーコードの詳細化指定をする場合

エラーコードの詳細化指定をする場合、SQL 連絡領域にメッセージが返却されるときは、HiRDB のメッセージ KFPA18040-E とともに出力されます。出力される形式は次のとおりです。なお、エラーコードの詳細化指定をするかどうかは、環境定義ファイルで指定できます。

```
KFPA18040-E (dddddd)aa....aa
```

dddddd : プラグインモジュール名

Text Search Plug-in では次のどれかが出力されます。

_phsgml : SGML プラグインで出力したメッセージ

_phngram : n-gram インデックスプラグインで出力したメッセージ

_phcfwd : FREEWORD プラグインで出力したメッセージ

_phxfwrdd : IXFREEWORD プラグインで出力したメッセージ

aa....aa : (ID=xxxxx) エラー内容

エラー内容が出力されます。

ID=xxxxx の形式で Text Search Plug-in でのメッセージ ID が出力されます。

メッセージ ID の意味については、「付録 F.6 SGML 文書の登録エラーメッセージ一覧」を参照してください。

(2) pdload で SGML 文書の登録エラーとなった場合

pdload で SGML 文書の登録エラーとなった場合、HiRDB のメッセージ KFPL31018-E とともにエラー情報ファイルに出力されます。出力される形式は次のとおりです。

```
KFPL31018-E PLUGIN function error occurred, line=aa....aa, reason="bb....bb"
```

aa....aa : 行番号

エラーが出力された行番号が出力されます。

bb....bb : (ID=xxxxx) エラー内容

エラー内容が出力されます。

ID=xxxxx の形式で Text Search Plug-in でのメッセージ ID が出力されます。

メッセージ ID の意味については、「付録 F.6 SGML 文書の登録エラーメッセージ一覧」を参照してください。

KFPL31018-E のメッセージの詳細については、マニュアル「HiRDB メッセージ」を参照してください。

付録 F.6 SGML 文書の登録エラーメッセージ一覧

SGML 文書の登録時に出力される，Text Search Plug-in でのメッセージ ID の一覧を次の表に示します。

表 F-2 SGML 文書の登録エラーメッセージ一覧

| メッセージ ID | エラー内容 |
|----------|---|
| 55002 | 必須の要素が指定されていません。要素の指定を確認してください。 |
| 55003 | 応用特有情報の指定はサポートしていません。無視します。 |
| 55004 | 要素 "%s" の属性 "%s" が二重に指定されています。重複する属性を削除してください。 |
| 55005 | マーク最小化は使われていないので、属性 "%s" の値は省略できません。 |
| 55006 | 要素 "%s" の属性 "%s" 中で未定義の一般実体 "%s" を参照しています。 |
| 55007 | 要素 "%s" の属性 "%s" 中で未定義の記法 "%s" を参照しています。 |
| 55008 | 宣言値が "CDATA" でない属性には、空の属性値は指定できません。属性値を指定してください。 |
| 55010 | "%s" の内容がありません。"%s" は必須要素のため、省略できません。内容が空の場合は開始タグを指定してください。 |
| 55012 | 要素型 "%s" の属性定義並びは定義済みです。 |
| 55013 | 記法 "%s" のデータ属性定義並びは定義済みです。 |
| 55014 | 文書型 "%s" は定義済みです。 |
| 55015 | 要素 "%s" は定義済みです。 |
| 55016 | 一般実体 "%s" は定義済みです。先に指定した一般実体を有効にして処理します。 |
| 55017 | 記法 "%s" は定義済みです。 |
| 55018 | 引数実体 "%s" は定義済みです。先に指定した引数実体を有効にして処理します。 |
| 55019 | 短縮参照対応表 "%s" は定義済みです。 |
| 55020 | 文書インスタンス中では、総合要素型の指定はできません。短縮参照使用宣言を確認してください。 |
| 55021 | 要素 "%s" には内容参照の属性が指定されているので、記法属性 "%s" は指定できません。 |
| 55023 | 文字番号 "%d" から "%d" までが重複して定義されています。文字集合のすべての文字番号を重複なく定義してください。 |
| 55024 | 文字番号 "%d" から "%d" までが定義されていません。文字集合のすべての文字番号を定義してください。 |
| 55026 | CONCUR 機能はサポートしていません。CONCUR NO が指定されたものと見なして処理します。 |
| 55027 | #CURRENT 属性を持つ要素は、最初の出現に限り、#CURRENT 属性の指定を省略することはできません。 |
| 55028 | DATATAG 機能はサポートしていません。 |

| メッセージ ID | エラー内容 |
|----------|--|
| 55029 | 既定実体は文書交換の際に使用できません。 |
| 55030 | LCNMCHAR と UCNMCHAR に指定された文字数が異っています。同一の文字数となるように指定してください。 |
| 55031 | LCNMSTRT と UCNMSTRT に指定された文字数が異っています。同一の文字数となるように指定してください。 |
| 55032 | ID 属性の属性名として既に "%s" が使用されています。ID 属性の属性名には同じ名前を指定してください。 |
| 55033 | 指定された文書型 "%s" は定義されていません。 |
| 55034 | データソーススタックが満杯です。処理を終了します。 |
| 55035 | 追加機能文字の名称 "%s" が重複して指定されています。 |
| 55036 | 属性定義並び内に属性名 "%s" が複数回指定されています。属性名は一つの属性定義並び内で重複して指定できません。 |
| 55037 | 容量 "%s" は既に設定されています。 |
| 55038 | 属性定義並び内に字句 "%s" を重複して指定することはできません。 |
| 55039 | 文字番号 "%d" が複数の機能文字に割り当てられています。 |
| 55040 | 属性定義並び内に ID 属性の宣言を複数指定することはできません。 |
| 55041 | 文字番号 %d (0x%x) の文字が、名前文字と名前開始文字に重複して指定されています。 |
| 55042 | 属性定義並び内に NOTATION 属性の宣言を複数指定することはできません。 |
| 55043 | 同一グループ内に字句 "%s" を重複して指定することはできません。 |
| 55044 | 属性が定義されていない要素 "%s" に属性が指定されています。属性の指定を削除してください。 |
| 55045 | 属性指定並びを指定しないときは、データ属性指定自体を指定することはできません。データ属性指定を削除してください。 |
| 55046 | 要素型 "%s" は内容が EMPTY なので、#CONREF 属性を持つことはできません。 |
| 55047 | 要素型 "%s" は内容が EMPTY なので、NOTATION 属性を持つことはできません。 |
| 55049 | 終了タグを指定できない要素 "%s" の終了タグが指定されています。終了タグを削除してください。 |
| 55050 | 開いていない要素 "%s" の終了タグが指定されています。終了タグを削除してください。 |
| 55051 | 文書型名指定を伴う実体参照はサポートしていません。無視します。 |
| 55052 | 未定義の記法名 "%s" をもつ実体が参照されています。 |
| 55053 | 文字データの途中で文書が終わっています。文字データの末尾を示す区切り子が欠けていないか確認してください。 |
| 55054 | コメントの途中で文書が終わっています。コメントの末尾を示す区切り子が欠けていないか確認してください。 |
| 55055 | 終了タグの途中で文書が終わっています。タグの末尾を示す区切り子が欠けていないか確認してください。 |

| メッセージ ID | エラー内容 |
|----------|---|
| 55056 | マーク付きセクションの途中で文書が終わっています。マーク付きセクションの末尾を示す区切り子が欠けていないか確認してください。 |
| 55057 | 引数データの途中で文書が終わっています。引数データの末尾を示す区切り子が欠けていないか確認してください。 |
| 55058 | 開始タグの途中で文書が終わっています。タグの末尾を示す区切り子が欠けていないか確認してください。 |
| 55059 | 文字記述の指定に誤りがあります。 |
| 55060 | 回避文字番号指定の指定に誤りがあります。 |
| 55061 | "%s" の長さが, "NAMELEN" に指定した制限値 "%d" を超えています。 |
| 55062 | 要素名, 実体名などの名前の文字数が制限 %d を超えています。 |
| 55063 | SGML テキストの要素のネストの深さが限界値 "%d" を超えています。処理を終了します。SGML テキストの要素のネストの深さ, 又は DTD 中の内容モデルを見直してください。 |
| 55064 | "TAGLVL" に指定した値が, システムの限界値 "%d" を超えています。"TAGLVL" の値を "%d" にして処理します。 |
| 55065 | 固定属性 "%s" にデフォルトと異なる値が指定されています。固定属性の属性値指定を削除してください。 |
| 55066 | FORMAL 機能はサポートしていません。指定は無視します。 |
| 55067 | 一般区切り子 "%s" に指定された文字列が機能文字を含んでいます。 |
| 55068 | 一般区切り子 "%s" に指定された文字列が機能文字だけから構成されています。 |
| 55069 | 指定された文字番号 "%d" は, 許容される最大値 "%d" を超えています。文字番号の値を確認してください。 |
| 55070 | 固有識別子 "%s" は他で既に指定されています。別の固有識別子を指定してください。 |
| 55071 | 参照されていない固有識別子 "%s" が指定されています。 |
| 55072 | 要素型 "%s" に使用する短縮参照対応表は指定済みです。先に指定した短縮参照対応表を有効にして処理します。 |
| 55073 | 有効文書型の指定 "%s" に誤りがあります。この指定は無視します。 |
| 55074 | 名前 "%s" は追加機能名称として指定することはできません。 |
| 55075 | 属性名 "%s" に誤りがあります。 |
| 55076 | 属性値 "%s" に誤りがあります。 |
| 55077 | 容量集合に指定した名前 "%s" に誤りがあります。 |
| 55078 | 容量値 "%s" に誤りがあります。 |
| 55079 | 最小データ中には許されない文字 (code=0x%x) が検出されました。 |
| 55080 | 内容字句の指定に誤りがあります。 |
| 55081 | 要素 "%s" の終了タグ内に不当なデータ "%s" があります。 |
| 55082 | 外部 DTD 中に不当なデータ "%s" があります。 |

| メッセージ ID | エラー内容 |
|----------|--|
| 55083 | 公開識別子定義ファイル中に不当なデータ "%s" があります。処理を終了します。 |
| 55084 | システム識別子定義ファイル中に不当なデータ "%s" があります。処理を終了します。 |
| 55085 | 要素 "%s" の開始タグ内に不当なデータ "%s" があります。 |
| 55086 | 属性の宣言値指定に誤りがあります。 |
| 55087 | 属性のデフォルト値指定に誤りがあります。 |
| 55088 | ID 属性のデフォルト値には、#IMPLIED 又は #REQUIRED 以外指定できません。 |
| 55089 | 要素型名の指定に誤りがあります。 |
| 55090 | 実体文の指定に誤りがあります。 |
| 55091 | 実体型の指定 "%s" に誤りがあります。 |
| 55092 | 一般区切り子に指定した名前 "%s" に誤りがあります。 |
| 55093 | 対応表指定に誤りがあります。 |
| 55094 | マークアップ宣言に誤りがあります。 |
| 55095 | 内容中に不当なマークアップ宣言があります。 |
| 55096 | 名前文字としては指定できない文字 (文字番号 %d (0x%x)) が指定されています。 |
| 55097 | 量集合に指定した名前 "%s" に誤りがあります。 |
| 55098 | 予約名使用に指定した名前 "%s" に誤りがあります。 |
| 55099 | 許される予約語 "%s" 以外の文字列が指定されています。 |
| 55100 | この位置には出現し得ない要素またはデータを検出しました。文書が DTD に従って正しく構成されているかどうかチェックしてください。 |
| 55101 | 接続子 and で指定されたモデル群に対応して必要となる内容がそろっていません。%d 番目の内容字句に対応する内容を補ってください。 |
| 55102 | 外部 DTD が途中で終わっています。処理を終了します。 |
| 55103 | TOTALCAP に指定した容量値 "%d" では不十分です。全容量の総和 "%d" 以上を指定してください。 |
| 55104 | 整数値演算中にオーバーフローが発生しました。 |
| 55105 | この位置に文字データを置くことはできません。 |
| 55106 | この位置にデータ "%s" を置くことはできません。 |
| 55107 | この位置に要素 "%s" を置くことはできません。 |
| 55108 | 機能種別の指定に誤りがあります。 |
| 55109 | LINK 機能はサポートしていません。連結型宣言の指定があると正しく解析できないことがあります。 |
| 55110 | 定義していない要素型 "%s" を短縮参照対応宣言に指定しています。 |
| 55111 | マーク認識抑止機能はサポートしていません。 |
| 55112 | 関連する要素型の指定がありません。 |

| メッセージ ID | エラー内容 |
|----------|--|
| 55113 | 属性定義に誤りがあります。属性名、宣言値及びデフォルト値の指定を確認してください。 |
| 55114 | 終了タグを省略できない要素 "%s" の終了タグがありません。 |
| 55115 | 終了タグ省略の指定がありません。省略可 (O) 又は省略不可 (-) を指定してください。 |
| 55116 | 開始タグを省略できない要素 "%s" の開始タグがありません。 |
| 55117 | 同一グループ内に複数種類の接続子を指定することはできません。 |
| 55118 | マーク区間外でマーク区間終了 "%s" を指定することはできません。 |
| 55119 | 結合記法名で指定した名前 "%s" は記法名として宣言されていません。名前の指定および記法宣言を確認してください。 |
| 55120 | 置換可能文字データ (RCDATA) 中で NDATA 実体への参照はできません。 |
| 55121 | 名前グループ内に名前以外のものが指定されています。 |
| 55122 | 名前字句グループ内に名前字句以外のものが指定されています。 |
| 55123 | 非 SGML 文字 %d (0x%x) を検出しました。 |
| 55124 | データ属性を持たない記法 "%s" に対して属性が指定されています。データ属性指定を確認してください。 |
| 55125 | 容量名でないもの "%s" が指定されています。 |
| 55126 | 引数表記の指定に短縮参照区切り子でないものが指定されています。 |
| 55127 | 引数実体には実体型 "CDATA" は指定できません。 |
| 55128 | 文書型 "%s" に対応する要素宣言がありません。文書型宣言で指定した文書型に対応する要素宣言を追加してください。 |
| 55130 | 引数実体には実体型 "NDATA" は指定できません。 |
| 55131 | 引数実体には実体型 "SDATA" は指定できません。 |
| 55133 | 引数実体には実体型 "SUBDOC" は指定できません。 |
| 55134 | 旧形式の規格バージョン "ISO 8879-1986" が指定されています。"ISO 8879:1986" を指定してください。 |
| 55135 | 接続子 or で指定されたモデル群に対応する要素がどれも指定されていません。要素の指定を確認してください。 |
| 55136 | 置換可能文字データ (RCDATA) 中で PI 実体への参照はできません。 |
| 55137 | 公開識別子 "%s" に対応するシステム識別子 "%s" と、指定されたシステム識別子 "%s" とが一致しません。公開識別子に対応するシステム識別子で処理します。外部識別子の指定を確認してください。 |
| 55138 | %d 個のエラーを検出しました。処理を終了します。 |
| 55139 | RANK 機能はサポートしていません。 |
| 55140 | 実体 "%s" が再帰的に参照されています。実体宣言を確認してください。 |
| 55141 | ID 参照値 "%s" は固有識別子として定義されていません。 |
| 55142 | 追加機能文字の文字番号が指定されていません。 |

| メッセージ ID | エラー内容 |
|----------|--|
| 55143 | 応用特有情報が指定されていません。 |
| 55144 | 短縮参照使用宣言が文書型宣言の中で現われるときは、統合要素型の指定を省略することはできません。 |
| 55145 | 要素型名または表記法名が指定されていません。 |
| 55146 | 必須の属性 "%s" の指定がありません。必須属性を指定してください。 |
| 55147 | 基本文字集合が指定されていません。 |
| 55148 | 見出し語 "SGMLREF" の後に名前が指定されていません。 |
| 55149 | 容量集合が指定されていません。 |
| 55150 | 名前の後に数が指定されていません。容量集合の指定を確認してください。 |
| 55151 | 文字記述が指定されていません。 |
| 55152 | 文字番号が指定されていません。公開具象構文の指定を確認してください。 |
| 55153 | 具象構文が指定されていません。 |
| 55154 | 具象構文範囲が指定されていません。 |
| 55155 | 要素内容の指定に誤りがあります。宣言内容又は内容モデルの指定を確認してください。 |
| 55156 | 区切り子 "%s" を指定すべき個所に文字列 "%s" があります。 |
| 55157 | 区切り子集合が指定されていません。 |
| 55158 | 名前の後に引数表記が指定されていません。一般区切り子の指定を確認してください。 |
| 55159 | 文字集合区分が指定されていません。 |
| 55160 | "=" が所定の位置に指定されてません。公開識別子定義ファイルの内容を確認してください。処理を終了します。 |
| 55161 | ファイルシステム種別が所定の位置に指定されてません。システム識別子定義ファイルの内容を確認してください。処理を終了します。 |
| 55162 | 文書型宣言が所定の位置にありません。文書型宣言の前にある不当なデータを削除してください。または、文書型宣言の指定を確認してください。 |
| 55163 | 文書型名の指定に誤りがあります。 |
| 55164 | 具象構文範囲に "DOCUMENT" または "INSTANCE" が指定されていません。 |
| 55165 | 文書文字集合が指定されていません。 |
| 55166 | 排除対象である要素が指定されています。要素の指定を確認してください。 |
| 55167 | 実体名の指定に誤りがあります。 |
| 55168 | 属性 "%s" に "ENTITY" 型以外の属性値が指定されています。 |
| 55169 | 属性 "%s" に "ENTITIES" 型以外の属性値が指定されています。 |
| 55170 | 機構使用が指定されていません。 |
| 55171 | 機能文字指定が指定されていません。 |
| 55172 | 一般区切り子が指定されていません。 |
| 55173 | 総称識別子(要素型名)の指定に誤りがあります。 |

| メッセージ ID | エラー内容 |
|----------|---|
| 55174 | 属性 "%s" に属性値集合で定義されていない属性値が指定されています。 |
| 55175 | 属性 "%s" に "IDREF" 型以外の属性値が指定されています。 |
| 55176 | 属性 "%s" に "IDREFS" 型以外の属性値が指定されています。 |
| 55177 | 属性 "%s" に "ID" 型以外の属性値が指定されています。 |
| 55178 | 見出し語 "LCNMCHAR" 又はこれに対する引数表記が命名方法の中に指定されていません。 |
| 55179 | 見出し語 "LCNMSTRT" 又はこれに対する引数表記が命名方法の中に指定されていません。 |
| 55180 | 対応表名の指定に誤りがあります。 |
| 55181 | 見出し語 "NAMECASE" が命名方法の中に指定されていません。 |
| 55182 | 見出し語 "ENTITY" 又はこれに対する "YES" か "NO" が命名方法の中に指定されていません。 |
| 55183 | 見出し語 "GENERAL" 又はこれに対する "YES" か "NO" が命名方法の中に指定されていません。 |
| 55184 | 属性 "%s" に "NAME" 型以外の属性値が指定されています。 |
| 55185 | 属性 "%s" に "NAMES" 型以外の属性値が指定されています。 |
| 55186 | 属性 "%s" に "NMTOKEN" 型以外の属性値が指定されています。 |
| 55187 | 属性 "%s" に "NMTOKENS" 型以外の属性値が指定されています。 |
| 55188 | 命名方法が指定されていません。 |
| 55189 | 記法名前群の指定に誤りがあります。 |
| 55190 | 記法識別子の指定に誤りがあります。 |
| 55191 | 記法名の指定に誤りがあります。 |
| 55192 | 属性 "%s" に "NOTATION" 型以外の属性値が指定されています。 |
| 55193 | 記法名又はその名前群が結合記法名に指定されていません。 |
| 55194 | 数値が指定されていません。 |
| 55195 | 属性 "%s" に "NUMBER" 型以外の属性値が指定されています。 |
| 55196 | 属性 "%s" に "NUMBERS" 型以外の属性値が指定されています。 |
| 55197 | 属性 "%s" に "NUTOKEN" 型以外の属性値が指定されています。 |
| 55198 | 属性 "%s" に "NUTOKENS" 型以外の属性値が指定されています。 |
| 55199 | タグ省略の指定がありません。省略可 (O) 又は省略不可 (-) を指定してください。 |
| 55200 | 実体文が引数表記になっていません。 |
| 55201 | 容量集合に "PUBLIC" 又は "SGMLREF" が指定されていません。 |
| 55202 | 公開識別子が指定されていません。 |
| 55203 | 量集合が指定されていません。 |

| メッセージ ID | エラー内容 |
|----------|--|
| 55204 | 見出し語 "RE" 又はこれに対する文字番号が機能文字指定の中に指定されていません。 |
| 55205 | 置き換える予約名が指定されていません。予約名使用の指定を確認してください。 |
| 55206 | 予約名使用が指定されていません。 |
| 55207 | 見出し語 "RS" 又はこれに対する文字番号が機能文字指定の中に指定されていません。 |
| 55208 | 見出し語 "SGMLREF" が指定されていません。 |
| 55209 | 短縮参照区切り子に "SGMLREF" 又は "NONE" が指定されていません。 |
| 55210 | 見出し語 "SPACE" 又はこれに対する文字番号が機能文字指定の中に指定されていません。 |
| 55211 | 短縮参照区切り子が指定されていません。 |
| 55212 | 引数表記の指定に誤りがあります。短縮参照区切り子又は対応実体名を指定してください。 |
| 55213 | SGML 規格のバージョンが指定されていません。"ISO 8879:1986" を指定してください。 |
| 55214 | #CURRENT 属性 "%s" を持つ要素 "%s" は、最初の出現に限り、開始タグを省略することはできません。 |
| 55215 | 見出し語 "%s" が指定されていません。 |
| 55216 | 具象構文に "PUBLIC" 又は "SHUNCHAR" が指定されていません。 |
| 55217 | 見出し語 "UCNMCHAR" 又はこれに対する引数表記が命名方法の中に指定されていません。命名方法の指定を確認してください。 |
| 55218 | 見出し語 "UCNMSTRT" 又はこれに対する引数表記が命名方法の中に指定されていません。命名方法の指定を確認してください。 |
| 55219 | SGML 宣言で "SHORTTAG YES" と指定されていないため、属性名及び値標識 "%s" を省略することはできません。 |
| 55220 | "YES" または "NO" が指定されていません。 |
| 55221 | 認識モードスタックが満杯です。処理を終了します。 |
| 55222 | 短縮参照対応表宣言内で短縮参照区切り子を重複して指定することはできません。 |
| 55223 | SGML 宣言で "OMITTAG NO" と指定されているので、開始タグ最小化には "-" を指定してください。 |
| 55224 | 要素内容が "EMPTY" のため、終了タグは指定できません。終了タグを省略可 (O) で処理します。 |
| 55225 | 要素 "%s" は宣言内容を持っているため、開始タグは省略できません。開始タグの省略不可 (-) を指定してください。 |
| 55226 | 要素 "%s" は必須属性を持っているため、開始タグは省略できません。開始タグの省略不可 (-) を指定してください。 |
| 55227 | 置換可能文字データ (RCDATA) 中で SGML 部分文書実体への参照はできません。 |
| 55228 | SUBDOC 機能は使用できません。 |

| メッセージ ID | エラー内容 |
|----------|---|
| 55229 | SUBDOC 機能はサポートしていません。SUBDOC NO が指定されたものと見なして処理します。 |
| 55230 | システム識別子が指定されていません。 |
| 55231 | 実体型が "SUBDOC" 又は "NDATA" である外部実体への参照はサポートしていません。また、置換可能引数データの中から外部実体への参照もサポートしていません。実体参照及び対応する実体宣言の指定を確認してください。 |
| 55232 | 定義されていない要素型名 "%s" が指定されています。 |
| 55233 | 一般実体 "%s" は定義されていません。 |
| 55234 | 定義されていない対応表名 "%s" が指定されています。 |
| 55235 | 引数実体 "%s" は定義されていません。 |
| 55236 | 区切り子の対応が不正です。 |
| 55237 | 参照される実体として完全ではありません。実体が外部 DTD ならば実体中の各宣言が完全かを確認してください。また、実体が引数実体ならば完全な引数の列かを確認してください。 |
| 55238 | SGML 文書の入力途中で終わりました。SGML 文書を確認してください。 |
| 55239 | 基本文字集合の公開識別子 "%s" に対応するシステム識別子は定義されていません。 |
| 55240 | 規格参照容量集合以外の公開識別子 "%s" が指定されています。指定した公開識別子はサポートしていません。容量集合の公開識別子を確認してください。 |
| 55241 | 具象構文の公開識別子に未知の値 "%s" が指定されています。デフォルトの具象構文で置き換えます。 |
| 55242 | ファイルシステム種別の指定 "%s" に誤りがあります。処理を終了します。 |
| 55243 | 文字参照の機能名 "%s" は定義されていません。定義されている機能名又は文字番号を指定してください。 |
| 55244 | 公開識別子 "%s" に対応するシステム識別子は定義されていません。 |
| 55245 | バージョンに "%s" が指定されています。"ISO 8879:1986" を指定してください。 |
| 55246 | DTD name unmatched.dtd name '%s' in document, '%s' in plugin description. |
| 55247 | %s errors occurred during parsing sgml document.file=%s line=%d |
| 55248 | External entity reference is not supported. name(%s) |
| 55249 | PUBLIC is not supported in SGML document/DTD. |
| 55300 | 最上位要素の終了以降に不当なデータ "%s" があります。不当なデータは無視して処理します。 |
| 55301 | 文字参照に指定した文字コードが限界値を超えています。文字コードの値を確認してください。 |
| 55302 | 実体参照のネスト数が %d を超えています。実体参照宣言を確認してください。 |
| 55303 | 不当なデータ "%s" があります。 |
| 55304 | この位置に要素 "%s" は指定できません。 |
| 56002 | 最上位要素の終了以降に不当なデータ "%s" があります。不当なデータは無視して処理します。 |

| メッセージ ID | エラー内容 |
|----------|---|
| 56004 | 入力文書に検索対象となる登録データがありません。 |
| 56005 | ESIS-B データ出力先ポインタを格納する領域へのアドレスに NULL が指定されています。 |
| 56006 | ESIS-B データサイズを格納する領域へのアドレスに NULL が指定されています。 |
| 56007 | 特定文字データ "%s" に対応するデータが正規化パラメタファイルに指定されていません。 |
| 56008 | 同じ構成要素が既に指定されています。 |
| 56009 | 指定方法に誤りがあります。 |
| 56010 | "%s" は既に指定されています。 |
| 56012 | 正規化できない要素型名を正規化パラメタファイルで指定しました。 |
| 56013 | プレーンテキスト文書のファイル名が指定されていません。 |
| 56014 | プレーンテキスト文書に挿入する文書型名が指定されていません。 |
| 56015 | プレーンテキスト文書に対応するシステム識別子が指定されていません。 |
| 56016 | プレーンテキスト文書ファイルのファイルサイズが 0 です。データが無いため、処理を中止します。 |

(凡例)

エラー内容中の "%s", "%d", "%x" には、直前の項目の値が設定されます。

付録 G 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。

DTD

SGML で文書を作成するときの、文書構造の定義です。

DTD ファイル

構造文書用の列に対して、登録する文書の構造を定義するためのファイルです。一つの列に一つの DTD ファイルを作成します。

FREWORD 型

文字列データを扱う抽象データ型のことです。

HiRDB 管理者

HiRDB の運用コマンドの実行者で、スーパーユーザが登録したシステム管理者用のユーザ ID でログインしたユーザのことです。HiRDB のディレクトリおよびファイルの所有者です。

IXFREWORD 型

FREWORD 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。

RD エリア

データベースの表、インデクス、およびデータディクショナリを格納するデータ領域のことです。

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。

SGML 定義情報

DTD と正規化パラメタのことです。SGML 文書の登録や検索に必要なこれらの情報をまとめて、SGML 定義情報と呼びます。

UCS2 (Universal multi-octet Character Set 2)

ISO によって規定された国際文字コード ISO/IEC 10646-1 の 16 ビット文字コードです。

UCS4 (Universal multi-octet Character Set 4)

ISO によって規定された国際文字コード ISO/IEC 10646-1 の 32 ビット文字コードです。

(ア行)

異表記展開

表記方法が異なるものを同じ意味の語とみなすことです。かたかな異表記、アルファベット異表記、全角半角異表記などがあります。

インクリメンタル

インデクス作成単位の文字数を増やすことです。

インデクス作成方法

登録している文書または文字列データに対してインデクスを作成する場合、次の方法があります。

- 即時登録
インデクス作成とインデクスへの文書登録を同時に実行する方法です。
- 初期分割登録または分割遅延登録
インデクスへ反映しないで、文書やデータの登録だけを実行し、インデクス作成開始ユーティリティを使ってインデクスへの文書登録を実行する方法です。

重み

検索条件を複数指定するときに、各検索条件に対して付ける重要度のことです。

(カ行)

概念検索

検索条件として指定した文章（種文章）をキーに、似た内容（概念）を持つ文書を検索する検索方法です。

近傍条件検索

単語と単語の間の文字数を条件として指定する検索方法です。

検索結果集合

検索条件に合致した文書の集合です。

検索条件

データベースの文書を検索するための条件です。

検索ターム

検索条件に指定するキーワードです。

検索用特徴ターム

種文章から抽出した特徴タームの中で、その種文章の概念を表すために採用されたタームのことです。種文章の特徴をより明確に示している特徴タームから順に、検索用特徴タームとして採用されます。

更新前ログ取得モード

UAPまたはユー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またはユーティリティが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す。

構造名

SGML 文書中のタグのことです。SGML 文書の内容は、このタグによって要素ごとに分けられています。

(サ行)

差分インデクス

文書の登録性能を向上させるために使用する小容量の一時的な登録用インデクスのことです。

除外文字検索

検索タームの一部または前後に特定の文字がある場合を除外して、該当する文書を検索することです。

スコア

検索条件に対して、どの程度の適応度があるかを示した得点です。

正規化パラメタ

正規化パラメタファイルの構成要素の一つです。SGML 文書中のタグの制御方法を定義します。

正規化パラメタファイル

SGML 文書内のタグや文書中の特定データを制御するパラメタを記述するファイルです。正規化パラメタファイルは「正規化パラメタ」と「特定文字データ変換マップ」で構成されます。

属性名

構造に付加された属性の名前のことです。

(タ行)

種文章

概念検索で、似通った文書を探す基になる文章のことです。

抽象データ型

ユーザが複雑な構造を持つデータとその操作を独自に定義し、利用できるデータ型のことです。

同義語

同じ意味を持つ言葉のことです。例えば、「コンピュータ」の同義語は「電子計算機」となります。

同義語・異表記展開

検索タームの同義語および異表記を呼び出すことです。同義語は同義語辞書から呼び出します。異表記は、システムで決められた規則に従って呼び出されます。

同義語辞書

同義語の展開方法を定義したファイルです。

登録・検索同時実行

排他制御の単位を小さくすることで、SQL によるデータ登録と検索の同時実行を可能とする機能で

す。

特徴ターム

概念検索で、種文章の内容（概念）を特徴付けているタームのことです。

特徴ターム抽出しきい値

種文章から特徴タームを抽出する際に、出現確率が高いタームを除くためのしきい値のことです。

特定文字データ変換マップ

正規化パラメタファイルの構成要素の一つです。SGML 文書中の特定の文字の変換方法を定義します。

(八行)

プラグインインデクスの遅延一括作成

プラグインインデクスを定義した表に行データを追加したとき、プラグインインデクスのデータ追加処理をしないで、データベース再編成ユティリティを使用して、あとで一括してプラグインインデクスのデータ追加処理をする機能です。

プラグインインデクスを定義した表の行データを大量追加（または大量更新）するときこの機能を使用できます。

文書

データベースに登録するテキスト形式のファイルです。

文書検索プラグイン

SGML 文書の検索や登録など、文書を扱うためのプラグインです。文書検索プラグインは、SGMLTEXT 型という抽象データ型を扱う SGML プラグインと、SGMLTEXT 型に対応する NGRAM 型というインデクス型を扱う n-gram インデクスプラグインで構成されています。

(マ行)

文字列検索プラグイン

短い文字列データの検索や登録など、文字列データを扱うためのプラグインです。文字列検索プラグインは、FREEWORD 型という抽象データ型を扱う FREEWORD プラグインと、FREEWORD 型に対応する IXFREEWORD 型というインデクス型を扱う IXFREEWORD プラグインで構成されています。

(ヤ行)

ユーザ LOB 用 RD エリア

文書や画像、音声などの長大な可変長データを格納するための RD エリアのことです。次に示すデータを、ユーザ LOB 用 RD エリアに格納する必要があります。

- BLOB 型を指定した列（BLOB 列）
- 抽象データ型内の、BLOB 型を指定した属性

- プラグインインデクス

ユーザ用 RD エリア

ユーザが作成する表とインデクスを格納するための RD エリアのことです。

ユニット

一つのサーバマシン内の HiRDB の動作環境のことです。

(ラ行)

レジストリ LOB 用 RD エリア

レジストリ情報を管理する表 (レジストリ管理表) を格納するための RD エリアです。レジストリ機能を使用する場合に必要です。ただし、プラグインの種類によっては、レジストリ機能を使用しないものがあります。登録されるデータの長さによって、レジストリ LOB 用 RD エリアに格納するかどうかをシステムが自動的に決定します。また、レジストリ管理表に情報を登録したりする、操作のストアドプロシジャもこの RD エリアに格納します。

レジストリ機能

データ操作時にプラグインが使用するためのプラグイン固有の情報を、HiRDB が保持する機能のことです。

レジストリ用 RD エリア

レジストリ情報を管理する表 (レジストリ管理表) を格納するための RD エリアです。レジストリ機能を使用する場合に必要です。ただし、プラグインの種類によっては、レジストリ機能を使用しないものがあります。

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

数字

1RD エリア当たりの排他制御用共用メモリ
サイズ〔IXFREEWORD プラグイン〕 132
1RD エリア当たりの排他制御用共用メモリ
サイズ〔n-gram インデクスプラグイン〕
102
1 件登録時最大メモリサイズ
〔IXFREEWORD プラグイン〕 130
1 件登録時最大メモリサイズ〔n-gram イン
デクスプラグイン〕 99
1 文字近傍検索実行可否〔n-gram インデク
スプラグイン〕 101
1 文字近傍検索実行可否〔IXFREEWORD
プラグイン〕 132

B

BLOB〔用語解説〕 385

C

concept_with_score 237
concept_with_terms 241
contains 227
contains_with_score 228
CREATE INDEX〔文書検索プラグイン〕 68
CREATE INDEX〔文字列検索プラグイン〕
106
CREATE TABLE〔文書検索プラグイン〕 65
CREATE TABLE〔文字列検索プラグイン〕
105

D

DTD 57
DTD〔用語解説〕 385
DTD の登録 63
DTD ファイル 57
DTD ファイル〔用語解説〕 385

E

extracts 232

F

FREEWORD 226
FREEWORD 型 4
FREEWORD 型〔用語解説〕 385
FREEWORD プラグイン 4

H

HiRDB 管理者〔用語解説〕 385
HiRDB を OS から削除する場合の注意事項
139

I

IXFREEWORD 型 5
IXFREEWORD 型〔用語解説〕 385
IXFREEWORD プラグイン 5

N

n-gram インデクスプラグイン 4
n-gram インデクスプラグインの環境設定 95
n-gram インデクスプラグインの環境定義
ファイルの定義例 97
NGRAM 型 4
NGRAM 型〔用語解説〕 385
NOT 検索 9

P

pload 141
PDPLUGINNSUB 76
pdrorg 141
phcsyndicbkup 198
phcsyndicconv 197
phcsyndicreg 194
phcsyndicrm 196
phnchgcon 191

phnconbuild 185
 phncond 178
 phnconget 181
 phnconrstr 187
 phndelay 173
 phnexpnd 168
 phnidxls 159
 phnincmt 163
 phnmerge 166
 phnmodidx 175
 phnstartidx 170
 phsregsetup 150
 phssgmlreg 151
 phssgmlrm 152
 phssgmlval 153
 phssyndicbkup 158
 phssyndicconv 157
 phssyndicreg 154
 phssyndicrm 156
 phxcond 216
 phxdelay 211
 phxexpnd 205
 phxidxls 199
 phxincmt 202
 phxmodidx 213
 phxstartidx 207
 PIC ファイル名 51

R

RD エリア〔用語解説〕 385
 RD エリア容量 31

S

SAMECASE 71, 108
 SAMEWIDE 71, 108
 score 230
 score_concept 238, 239
 SGMLTEXT 224
 SGMLTEXT 型 4, 56
 SGMLTEXT 型〔用語解説〕 385
 SGML 出力 233
 SGML 定義情報 56

SGML 定義情報〔用語解説〕 385
 SGML 定義情報削除ユティリティ〔文書検索プラグイン〕 152
 SGML 定義情報参照ユティリティ〔文書検索プラグイン〕 153
 SGML 定義情報登録ユティリティ〔文書検索プラグイン〕 151
 SGML 定義情報の削除 91
 SGML 定義情報の作成 56
 SGML 定義情報の参照 91
 SGML 定義情報の登録 62
 SGML 定義情報の変更 90
 SGML 定義情報のメンテナンス 90
 SGML プラグイン 4
 SGML 文書中の改行コードの変換方法 59
 SGML 文書の正規化の例 61
 SGML 文書の制限 58
 SGML 文書の登録エラーメッセージ 373
 SGML 文書の登録エラーメッセージ一覧 375
 SQL 実行時の排他制御のモード（SQL に対応した論理ファイルの使用）〔FREEWORD プラグインおよび IXFREEWORD プラグイン〕 320
 SQL 実行時の排他制御のモード（SQL に対応した論理ファイルの使用）〔SGML プラグインおよび n-gram インデクスプラグイン〕 318
 SQL 連絡領域に出力されるメッセージ 325, 328

T

terms_concept 243
 Text Search Enhancer for English 5
 Text Search Plug-in Conceptual Extension 5, 9
 Text Search Plug-in システムの構成例 5
 Text Search Plug-in システムの構築 17
 Text Search Plug-in システムの構築の流れ 18
 Text Search Plug-in とは 2
 Text Search Plug-in のアンセットアップ 138

Text Search Plug-in の環境設定 49
 Text Search Plug-in の機能 2
 Text Search Plug-in の削除 136
 Text Search Plug-in の設定解除 135
 Text Search Plug-in のセットアップ (UNIX の場合) 49
 Text Search Plug-in の抽象データ型関数 221
 Text Search Plug-in の登録 50, 51
 Text Search Plug-in の登録運用の長所および短所 291
 Text Search Plug-in の特長 2
 Text Search Plug-in のバージョンアップ 51
 Text Search Plug-in のユティリティ 143, 144
 Text Search Plug-in を構成するプログラム 4
 Text Search Plug-in を利用した検索 7

U

UAP の更新およびバッチ登録による運用 293
 UCS2(Universal multi-octet Character Set 2)〔用語解説〕 385
 UCS2 から UCS4 への移行 81
 UCS4(Universal multi-octet Character Set 4)〔用語解説〕 385
 UCS4 のインデクス定義〔文字列検索プラグイン〕 115
 UCS4 のインデクス定義〔文書検索プラグイン〕 81
 utf-8 で使用できる文字コード 305

X

XML 出力 235

い

一括登録時最大メモリサイズ〔IXFREEWORD プラグイン〕 130
 一括登録時最大メモリサイズ〔n-gram インデクスプラグイン〕 99
 異表記オプション 254

異表記展開〔用語解説〕 386
 異表記展開検索〔概念検索〕 12
 異表記展開検索〔全文検索〕 8
 異表記展開検索時の部分展開検索モード〔IXFREEWORD プラグイン〕 131
 異表記展開指定〔概念検索条件式〕 281
 異表記展開指定〔文字列検索プラグイン〕 269
 異表記展開指定〔文書検索プラグイン〕 254
 異表記部分展開検索モード〔n-gram インデクスプラグイン〕 100
 異表記展開指定の形式〔概念検索条件式〕 281
 異表記展開指定の形式〔文字列検索プラグイン〕 269
 異表記展開指定の形式〔文書検索プラグイン〕 254
 異表記展開の例 257
 インクリメンタル 163, 202
 インクリメンタル〔用語解説〕 386
 インデクス型 4, 5
 インデクス作成開始ユティリティ〔文書検索プラグイン〕 170
 インデクス作成開始ユティリティ〔文字列検索プラグイン〕 207
 インデクス作成方法〔用語解説〕 386
 インデクス作成方法指定〔文書検索プラグイン〕 77
 インデクス作成方法指定〔文字列検索プラグイン〕 112
 インデクス作成方法指定の定義方法〔文書検索プラグイン〕 77
 インデクス作成方法指定の定義方法〔文字列検索プラグイン〕 112
 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 294
 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 295
 インデクス作成ルール追加機能 14, 20
 インデクス作成ルール追加機能使用時 (INDEX_ALL) のインデクス定義例 71

インデクス作成ルール追加機能使用時
 (INDEX_MEDIUM) のインデクス定義例
 71
 インデクス作成ルール追加機能のインデクス
 定義 70
 インデクス情報の取得ユティリティ〔文書検
 索プラグイン〕159
 インデクス情報の取得ユティリティ〔文字列
 検索プラグイン〕199
 インデクス情報変更ユティリティ〔文書検索
 プラグイン〕175
 インデクス情報変更ユティリティ〔文字列検
 索プラグイン〕213
 インデクス詰め替えユティリティ〔文書検索
 プラグイン〕178
 インデクス詰め替えユティリティ〔文字列検
 索プラグイン〕216
 インデクス定義時の注意 81
 インデクス定義ファイル〔バックアップ〕
 300
 インデクス定義ファイル〔系切り替え〕323
 インデクスのインクリメンタルユティリティ
 〔文書検索プラグイン〕163
 インデクスのインクリメンタルユティリティ
 〔文字列検索プラグイン〕202
 インデクスの拡張ユティリティ〔文書検索プ
 ラグイン〕168
 インデクスの拡張ユティリティ〔文字列検索
 プラグイン〕205
 インデクスの格納ページ数の設定項目と設定
 値 36,40
 インデクスの定義〔文書検索プラグイン〕67
 インデクスの定義〔文字列検索プラグイン〕
 106
 インデクスの定義方法 81
 インデクスのマージユティリティ〔文書検索
 プラグイン〕166
 インデクスファイル自動拡張機能のインデク
 ス定義〔文字列検索プラグイン〕107
 インデクスファイル自動拡張機能のインデク
 ス定義〔文書検索プラグイン〕69
 インデクス分割遅延登録運用変更ユティリ
 ティ〔文書検索プラグイン〕173

インデクス分割遅延登録運用変更ユティリ
 ティ〔文字列検索プラグイン〕211
 インデクス用グローバルバッファ最適化モー
 ド指定〔IXFREEWORD プラグイン〕133
 インデクス用グローバルバッファ最適化モー
 ド指定〔n-gram インデクスプラグイン〕
 102

え

英文検索機能のインデクス定義〔文字列検索
 プラグイン〕115
 英文検索機能のインデクス定義〔文書検索プ
 ラグイン〕81
 英文検索のインデクスに指定できる項目 82
 英文で登録する文書(テキストデータ)およ
 び文字列データに使用できる文字 304
 エラーコード詳細化指定〔FREEWORD プ
 ラグイン〕126
 エラーコード詳細化指定〔SGML プラグイ
 ン〕95
 エラーメッセージ 325

お

重み 9
 重み〔概念検索条件式〕283
 重み〔文書検索プラグイン〕260
 重み〔用語解説〕386
 重み指定の形式 260,283
 重み付き検索 9

か

概念検索 9
 概念検索〔用語解説〕386
 概念検索機能使用時のインデクス定義例 69
 概念検索条件式の形式 276
 概念検索条件式の指定例 286
 概念検索条件式の文法 276
 概念検索の特長 11
 概念検索用統計情報サイズ(1RD エリア当
 たりの phnconget 結果)43
 概念検索用統計情報の収集,統合および反映
 302

概念検索用統合統計情報サイズ
(phnconbuild 結果) 43
 環境情報の定義方法〔FREEWORD プラグ
イン〕 124
 環境情報の定義方法〔IXFREEWORD プラ
グイン〕 126
 環境情報の定義方法〔n-gram インデクスブ
ラグイン〕 95
 環境情報の定義方法〔SGML プラグイン〕
92
 環境定義ファイル〔SGML プラグイン〕 92
 環境定義ファイル〔バックアップ〕 301
 環境定義ファイル〔系切り替え〕 324
 環境定義ファイルに設定できる定義項目
〔FREEWORD プラグイン〕 125
 環境定義ファイルに設定できる定義項目
〔IXFREEWORD プラグイン〕 128
 環境定義ファイルに設定できる定義項目〔n-
gram インデクスプラグイン〕 97
 環境定義ファイルに設定できる定義項目
〔SGML プラグイン〕 93

き

距離 8, 258, 272
 近傍条件検索 8
 近傍条件検索〔用語解説〕 386
 近傍条件指定〔文字列検索プラグイン〕 272
 近傍条件指定〔文書検索プラグイン〕 258
 近傍条件指定の形式〔文字列検索プラグ
イン〕 273
 近傍条件指定の形式〔文書検索プラグイン〕
259

く

グローバルバッファ管理の運用 298

け

系切り替え 323
 検索因子 250, 267, 279
 検索因子の形式 251, 267, 279
 検索オプション 285

検索オプション〔IXFREEWORD プラグ
イン〕 131
 検索オプション〔n-gram インデクスプラ
グイン〕 99
 検索結果集合〔用語解説〕 386
 検索時最大ワークメモリサイズ
〔IXFREEWORD プラグイン〕 130
 検索時最大ワークメモリサイズ〔n-gram イ
ンデクスプラグイン〕 99
 検索条件〔用語解説〕 386
 検索条件式の指定に関する注意 246
 検索条件の AND/OR 論理条件 266
 検索ターム 7
 検索ターム〔用語解説〕 386
 検索タームによる絞り込み検索 13
 検索用特徴ターム 10
 検索用特徴ターム〔用語解説〕 386
 検索用特徴ターム出力 13
 検索用特徴ターム数〔n-gram インデクスブ
ラグイン〕 101
 検索用特徴タームを指定した検索 13
 検索用ワークファイルディレクトリ
〔IXFREEWORD プラグイン〕 130
 検索用ワークファイルディレクトリ〔n-
gram インデクスプラグイン〕 99
 検索抑止インデクスサイズ〔n-gram イン
デクスプラグイン〕 101
 検索抑止インデクスサイズ〔IXFREEWORD
プラグイン〕 132
 原文書出力 (FREEWORD 型の場合) 233
 原文書出力 (SGMLTEXT 型の場合) 232
 原文書データ保持モード指定〔SGML プラ
グイン〕 94

こ

更新前ログ取得モード〔用語解説〕 386
 更新前ログ取得モード 299
 構造検索パターン 247
 構造検索パターンの形式 248, 277
 構造名〔用語解説〕 387
 構造名指定の形式 249, 278
 構造名の指定〔概念検索条件式〕 277
 構造名の指定〔文書検索プラグイン〕 248

構造名のワイルドカード指定〔概念検索条件式〕 277
 構造名のワイルドカード指定〔文書検索プラグイン〕 249
 構造名を指定した検索〔概念検索〕 13
 構造名を指定した検索〔全文検索〕 9
 構造名を指定した検索条件のAND/OR 論理条件〔概念検索条件式〕 276
 構造名を指定した検索条件のAND/OR 論理条件〔文書検索プラグイン〕 247
 コンストラクタ関数の情報 141

さ

サーバ単位〔FREEWORD プラグイン〕 124
 サーバ単位〔IXFREEWORD プラグイン〕 127
 サーバ単位〔n-gram インデックスプラグイン〕 96
 サーバ単位〔SGML プラグイン〕 92
 削除対象要素リスト 59
 削除文字指定機能 15, 21
 削除文字指定機能使用時のインデクス定義例〔文書検索プラグイン〕 74
 削除文字指定機能使用時のインデクス定義例〔文字列検索プラグイン〕 110
 削除文字指定機能のインデクス定義〔文書検索プラグイン〕 74
 削除文字指定機能のインデクス定義〔文字列検索プラグイン〕 110
 差分インデクス 21, 76
 差分インデクス〔用語解説〕 387
 差分インデクス使用時のインデクス定義例 76
 差分インデクス定義 76
 差分インデクスによる運用 293
 差分インデクスの定義方法 76

し

資源の排他 318
 システム共通定義の変更 53, 139
 システム単位〔FREEWORD プラグイン〕 124

システム単位〔IXFREEWORD プラグイン〕 126
 システム単位〔n-gram インデックスプラグイン〕 95
 システム単位〔SGML プラグイン〕 92
 システムの制限 134
 指定できるワイルドカードおよび特殊文字 252
 絞込み検索でのリスト機能拡張 142
 絞り込み指定 285
 障害発生時の運用方法 322
 除外文字検索 252, 268
 除外文字検索〔用語解説〕 387
 初期分割登録 21
 初期分割登録〔文書検索プラグイン〕 77
 初期分割登録〔文字列検索プラグイン〕 112
 初期分割登録使用時のインデクス定義例〔文書検索プラグイン概念検索用インデクス〕 78
 初期分割登録使用時のインデクス定義例〔文書検索プラグイン全文検索用インデクス〕 78
 初期分割登録使用時のインデクス定義例〔文字列検索プラグイン〕 112

す

推奨する運用方法 289
 推奨する組み合わせ 296
 推奨する登録運用 293
 スコア〔用語解説〕 387
 スコアオプション 283
 スコアオプション指定の形式 284
 スコア検索 13
 スコア算出法指定〔IXFREEWORD インデクスプラグイン〕 131
 スコア算出法指定〔n-gram インデックスプラグイン〕 100
 スコア正規化指定 284
 スコア値 230, 239
 スコア有効値指定 284

せ

正規化パラメタ 59
 正規化パラメタ〔用語解説〕 387

正規化パラメタの指定方法 60
 正規化パラメタの登録 63
 正規化パラメタファイル 59
 正規化パラメタファイル〔用語解説〕 387
 正規化パラメタファイルの記述例 60
 性能向上の運用 298
 接続対象要素リスト 59
 設定できる項目〔n-gram インデクスプラグイン〕 97
 設定できる項目〔SGML プラグイン〕 93
 全角文字半角文字統一機能使用時のインデクス定義例〔文書検索プラグイン〕 72
 全角文字半角文字統一機能使用時のインデクス定義例〔文字列検索プラグイン〕 108
 全文検索 7
 全文検索機能使用時のインデクス定義例〔文書検索プラグイン〕 69
 全文検索機能使用時のインデクス定義例〔文字列検索プラグイン〕 107
 全文検索条件式の形式〔文書検索プラグイン〕 247
 全文検索条件式の形式〔文字列検索プラグイン〕 266
 全文検索条件式の指定例〔文書検索プラグイン〕 263
 全文検索条件式の指定例〔文字列検索プラグイン〕 274
 全文検索条件式の文法（文書検索プラグインの場合） 247
 全文検索条件式の文法（文字列検索プラグインの場合） 266
 全文検索でのスコア値取得 142
 全文検索の特長 8

そ

即時登録 21
 即時登録〔文書検索プラグイン〕 77
 即時登録〔文字列検索プラグイン〕 112
 属性値指定 261
 属性値指定の形式 261
 属性名〔用語解説〕 387

た

大小文字・全角文字半角文字統一機能 14, 20
 大小文字・全角文字半角文字統一機能のインデクス定義〔文書検索プラグイン〕 71
 大小文字・全角文字半角文字統一機能のインデクス定義〔文字列検索プラグイン〕 107
 大小文字統一機能, および全角文字半角文字統一機能使用時のインデクス定義例〔文書検索プラグイン〕 72
 大小文字統一機能, および全角文字半角文字統一機能使用時のインデクス定義例〔文字列検索プラグイン〕 108
 大小文字統一機能使用時のインデクス定義例〔文書検索プラグイン〕 72
 大小文字統一機能使用時のインデクス定義例〔文字列検索プラグイン〕 108
 代表スコアオプション指定 284
 濁音統一機能 15, 20
 濁音統一機能使用時のインデクス定義例 74
 濁音統一機能のインデクス定義〔文書検索プラグイン〕 73
 濁音統一機能のインデクス定義〔文字列検索プラグイン〕 109
 種文章 12, 10, 280
 種文章〔用語解説〕 387
 単純文字列指定〔概念検索条件式〕 280
 単純文字列指定〔文字列検索プラグイン〕 268
 単純文字列指定〔文書検索プラグイン〕 251
 単純文字列指定の形式 280
 単純文字列指定の形式〔文字列検索プラグイン〕 268
 単純文字列指定の形式〔文書検索プラグイン〕 251

ち

抽象データ型 4
 抽象データ型〔用語解説〕 387
 抽象データ型関数 220
 抽象データ型関数の文法 224

抽象データ型関数を使用した文書または文字列データの操作 219
 抽象データ型の選択 222
 抽象データ型の相違点 222

つ

通常のインデクス定義〔文書検索プラグイン〕 68
 通常のインデクス定義〔文字列検索プラグイン〕 106

て

データ格納領域の確保〔文書検索プラグイン〕 63
 データ格納領域の確保〔文字列検索プラグイン〕 105
 データディクショナリの RD エリア容量〔文字列検索プラグイン〕 39
 データディクショナリの RD エリア容量〔文書検索プラグイン〕 35
 データの登録〔文書検索プラグイン〕 85
 データの登録〔文字列検索プラグイン〕 120
 データベース再編成ユティリティ 141
 データベース作成の流れ〔文書検索プラグイン〕 55
 データベース作成の流れ〔文字列検索プラグイン〕 104
 データベース作成ユティリティ 141
 データベースの作成〔文書検索プラグイン〕 55
 データベースの作成〔文字列検索プラグイン〕 104
 テキスト検索条件の形式 250, 267, 279
 テキスト検索条件の指定 267
 テキスト検索条件の指定〔概念検索条件式〕 279
 テキスト検索条件の指定〔文書検索プラグイン〕 250
 デッドロックの防止 321

と

統一インデクスによる運用 296

同義語〔用語解説〕 387
 同義語・異表記展開〔用語解説〕 387
 同義語異表記展開用最大メモリサイズ〔n-gram インデクスプラグイン〕 100
 同義語異表記展開用最大メモリサイズ〔IXFREEWORD プラグイン〕 131
 同義語グループ内での文字種の制限 89
 同義語辞書 88
 同義語辞書〔バックアップ〕 300
 同義語辞書〔系切り替え〕 323
 同義語辞書〔用語解説〕 387
 同義語辞書移行ユティリティ〔文書検索プラグイン〕 157
 同義語辞書移行ユティリティ〔文字列検索プラグイン〕 197
 同義語辞書削除ユティリティ〔文書検索プラグイン〕 156
 同義語辞書削除ユティリティ〔文字列検索プラグイン〕 196
 同義語辞書テキストファイル 88
 同義語辞書テキストファイルの記述例 89
 同義語辞書登録ユティリティ〔文書検索プラグイン〕 154
 同義語辞書登録ユティリティ〔文字列検索プラグイン〕 194
 同義語辞書の作成〔文書検索プラグイン〕 88
 同義語辞書の作成〔文字列検索プラグイン〕 123
 同義語辞書の登録〔文書検索プラグイン〕 90
 同義語辞書の登録〔文字列検索プラグイン〕 123
 同義語辞書バックアップユティリティ〔文書検索プラグイン〕 158
 同義語辞書バックアップユティリティ〔文字列検索プラグイン〕 198
 同義語辞書ファイル〔FREEWORD プラグインおよびIXFREEWORD プラグイン〕 44
 同義語辞書ファイル〔SGML プラグインおよび n-gram インデクスプラグイン〕 41
 同義語辞書名に指定できない記号および英数字 155
 同義語展開検索〔概念検索〕 12
 同義語展開検索〔全文検索〕 8

同義語展開指定〔概念検索条件式〕 282
 同義語展開指定〔文字列検索プラグイン〕 272
 同義語展開指定〔文書検索プラグイン〕 257
 同義語展開指定の形式 282
 同義語展開指定の形式〔文字列検索プラグイン〕 272
 同義語展開指定の形式〔文書検索プラグイン〕 258
 統計情報収集ユティリティ〔文書検索プラグイン〕 181
 統計情報統合ユティリティ〔文書検索プラグイン〕 185
 統計情報反映ユティリティ〔文書検索プラグイン〕 187
 登録 51
 登録・検索同時実行〔用語解説〕 387
 登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行指定〔FREEWORD プラグイン〕 125
 登録・検索同時実行, または分割遅延登録・更新系 SQL 同時実行指定〔SGML プラグイン〕 94
 登録・検索同時実行〔文書検索プラグイン〕 86
 登録・検索同時実行〔文字列検索プラグイン〕 121
 登録・検索同時実行による運用 294
 登録・検索同時実行による運用の長所および短所 292
 登録運用で推奨するパラメタの組み合わせ 297
 登録運用の種類 290
 登録運用のログ取得方式の設定 299
 登録件数指定 119
 登録先インデクスの指定方法 76
 登録する文書(テキストデータ)および文字列データに使用できる文字 304
 登録文書サイズ上限値指定〔SGML プラグイン〕 94
 特殊文字の指定〔文字列検索プラグイン〕 274
 特殊文字の指定〔文書検索プラグイン〕 262

特徴ターム 10
 特徴ターム〔用語解説〕 388
 特徴ターム取得 RD エリア指定ユティリティ〔文書検索プラグイン〕 191
 特徴ターム抽出しきい値〔用語解説〕 388
 特徴ターム抽出しきい値〔n-gram インデックスプラグイン〕 101
 特定構造検索指定 261
 特定構造検索指定の形式 261
 特定構造指定の形式 262
 特定文字データ変換マップ 59, 60
 特定文字データ変換マップ〔用語解説〕 388

は

排他取得内容 318
 バックアップ取得方法 299

ひ

非構造文書を扱う場合に推奨する抽象データ型 223
 標準エラー出力に出力されるメッセージ 360
 表の格納ページ数の設定項目と設定値 35, 40
 表の削除 136
 表の作成〔文書検索プラグイン〕 65
 表の作成〔文字列検索プラグイン〕 105

ふ

ファイル容量 41
 複合条件検索 9
 プラグインインストールディレクトリ 50
 プラグインインデクスの遅延一括作成〔用語解説〕 388
 プラグインインデクスの遅延一括作成による運用 293
 プラグイン定義ファイル名 51
 プラグイン提供関数からの一括取得処理 142
 プラグイン名 49
 プレーン出力 234
 分割遅延登録 22
 分割遅延登録・更新系 SQL 同時実行使用時のインデクス定義例 80

分割遅延登録・更新系 SQL 同時実行の設定
88, 122

分割遅延登録〔文書検索プラグイン〕 77, 78

分割遅延登録〔文字列検索プラグイン〕 112

分割遅延登録使用時のインデクス定義例〔文
書検索プラグイン概念検索用インデクス〕 79

分割遅延登録使用時のインデクス定義例〔文
書検索プラグイン全文検索用インデクス〕 79

分割遅延登録使用時のインデクス定義例〔文
字列検索プラグイン〕 113

分割遅延登録への移行時に実行する、インデ
クス情報変更ユティリティの指定例〔文書検
索プラグイン〕 79

分割遅延登録への移行時に実行する、インデ
クス情報変更ユティリティの指定例〔文字列
検索プラグイン〕 113

分割遅延登録用ファイルオーバ時の処理
〔IXFREEWORD プラグイン〕 133

分割遅延登録用ファイルオーバ時の処理〔n-
gram インデクスプラグイン〕 102

文書〔用語解説〕 388

文書検索 2

文書検索プラグイン 4

文書検索プラグイン〔用語解説〕 388

め

メッセージ一覧 328

メッセージの記述形式 327

メッセージログ・標準エラー出力に出力され
るメッセージ 327

メッセージログに出力されるメッセージ 349

メモリ所要量 26

も

文字コードによる機能差異 22

文字種 89

文字単位インデクス作成機能使用時のインデ
クス定義〔文字列検索プラグイン〕 111

文字単位インデクス作成機能使用時のインデ
クス定義〔文書検索プラグイン〕 75

文字列検索 2

文字列検索プラグイン 4

文字列検索プラグイン〔用語解説〕 388

ゆ

有効種文章サイズ〔n-gram インデクスプラ
グイン〕 102

有効範囲〔n-gram インデクスプラグイン〕
103

有効範囲〔SGML プラグイン〕 95

ユーザ LOB 用 RD エリア〔用語解説〕 388

ユーザ LOB 用 RD エリア容量〔文字列検索
プラグイン〕 38

ユーザ LOB 用 RD エリア容量〔文書検索プ
ラグイン〕 32

ユーザ環境設定〔文書検索プラグイン〕 92

ユーザ環境設定〔文字列検索プラグイン〕
124

ユーザ用 RD エリア〔用語解説〕 389

ユーザ用 RD エリア容量〔文字列検索プラ
グイン〕 37

ユーザ用 RD エリア容量〔文書検索プラ
グイン〕 32

ユティリティ〔文書検索プラグイン〕 150

ユティリティ〔文字列検索プラグイン〕 194

ユティリティコマンドの排他制御のモード

〔FREEWORD プラグインおよび

IXFREEWORD プラグイン〕 320

ユティリティコマンドの排他制御のモード

(SGML プラグインおよび n-gram インデ
クスプラグイン) 319

ユティリティの記述形式 147

ユティリティの使用方法 144

ユニット〔用語解説〕 389

ユニット単位〔FREEWORD プラグイン〕
124

ユニット単位〔IXFREEWORD プラグイン〕
127

ユニット単位〔n-gram インデクスプラ
グイン〕 95

ユニット単位〔SGML プラグイン〕 92

よ

拗音統一機能 14, 20

拗音統一機能使用時のインデクス定義例〔文書検索プラグイン〕 73
 拗音統一機能使用時のインデクス定義例〔文字列検索プラグイン〕 109
 拗音統一機能のインデクス定義〔文書検索プラグイン〕 72
 拗音統一機能のインデクス定義〔文字列検索プラグイン〕 108

り

リソースの見積もり 26

れ

レジストリ LOB 用 RD エリア〔用語解説〕 389
 レジストリ LOB 用 RD エリア容量〔文字列検索プラグイン〕 41
 レジストリ LOB 用 RD エリア容量〔文書検索プラグイン〕 37
 レジストリ管理表のインデクス格納ページ数 37
 レジストリ管理表の格納ページ数 36
 レジストリ機能〔用語解説〕 389
 レジストリ機能の初期設定 52
 レジストリ情報の削除（SGMLTEXT 型の列を定義していた場合） 136
 レジストリ初期化ユティリティ〔文書検索プラグイン〕 150
 レジストリの初期化 63
 レジストリの排他 321
 レジストリ用 RD エリア〔用語解説〕 389
 レジストリ用 RD エリア容量〔文字列検索プラグイン〕 41
 レジストリ用 RD エリア容量〔文書検索プラグイン〕 36

ろ

ログ取得モード 299
 ログ取得モード〔用語解説〕 389
 ログ量〔FREEWORD プラグインおよび IXFREEWORD プラグイン〕 46

ログ量〔SGML プラグインおよび n-gram インデクスプラグイン〕 43
 ログレスモード 299
 ログレスモード〔用語解説〕 389
 論理演算検索〔概念検索〕 12
 論理演算検索〔全文検索〕 8
 論理積 8
 論理ファイル数 47
 論理和 8

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。

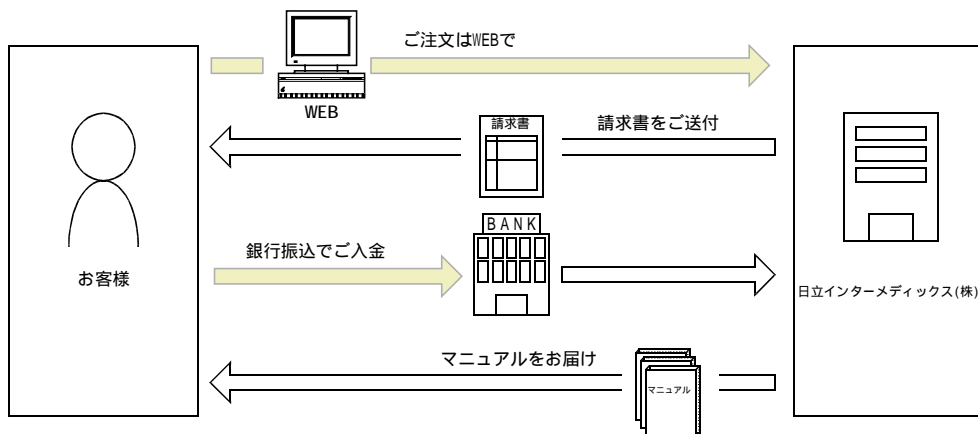
(1) マニュアル情報ホームページ「オンラインマニュアル」での公開

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。