Contents

About This Book ........................................................................................................................................... ix

CHAPTER 1  Introduction ................................................................................................................................. 1
   Capabilities of the Enhanced Full-Text Search Engine .......... 1
   High availability ................................................................. 3

CHAPTER 2  Understanding the Enhanced Full-Text Search Engine ....................................................... 5
   Components of the Enhanced Full-Text Search engine ...... 5
      The source table ............................................................ 5
      The Verity collections .................................................... 6
      Filters ........................................................................... 6
      The text_db database .................................................... 6
      The index table .............................................................. 7
      The text_events table ................................................... 8
      Relationships between the components ......................... 9
   How a full-text search works ............................................. 10

CHAPTER 3  Configuring Adaptive Server for Full-Text Searches ............................................................ 13
   Configuring Adaptive Server for an Enhanced Full-Text Search engine 13
      Enabling configuration parameters ................................. 14
      Running the installtextserver script ............................... 14
      Running the installmessages script ................................. 16
      Running the installevent script .................................... 17
      Naming the local Adaptive Server .................................. 18
   Creating and maintaining text indexes ............................... 19
      Setting up source tables for indexing ............................. 19
      Creating the text index and index table .......................... 20
      Granting permissions on text index proxy tables ............. 22
      Bringing the database online for full-text searches .......... 22
      Propagating changes to the text index ............................ 23
      Replicating text indexes ................................................ 24
      Example: enabling a new database for text searches ....... 24
Contents

Indexing the euro symbol ........................................................ 26

CHAPTER 4 Setting Up Verity Functions .......................................................... 27
Enabling query-by-example, summarization, and clustering........ 27
Editing the master style.prm file .................................................. 28
Editing individual style.prm files .............................................. 29
Setting up a column to use as a sort specification ....................... 30
Using filters on text that contains tags ....................................... 32
Creating a customthesaurus .................................................... 34
Examining the default thesaurus (optional) .............................. 35
Creating the control file ......................................................... 35
Creating the thesaurus ......................................................... 37
Replacing the defaultthesaurus with the customthesaurus .......... 37
Creating topics ......................................................................... 38
Creating an outline file ............................................................ 39
Creating a topic set directory ................................................... 40
Creating a knowledge base map .............................................. 40
Defining the location of the knowledge base map ................. 41
Executing queries against defined topics ................................. 41
Troubleshooting topics ........................................................... 42

CHAPTER 5 Writing Full-Text Search Queries ................................................. 43
Components of a full-text search query ...................................... 43
Default behaviour .................................................................. 44
Pseudo columns in the index table .......................................... 44
Using the score column to relevance-rank search results ....... 45
Using the sort_by column to specify a sort order .................... 46
Using the summary column to summarize documents ............ 47
Using pseudo columns to request clusteredresult sets .......... 48
Full-text search operators ....................................................... 50
Considerations when using Verity operators ........................... 51
Using the Verity operators ...................................................... 52
Operator modifiers .................................................................. 60

CHAPTER 6 System Administration ......................................................... 63
Starting the Enhanced Full-Text Search engine on UNIX ....... 63
Creating the runserver file ...................................................... 63
Starting the Enhanced Full-Text Search engine on Windows NT.. 65
Starting the Enhanced Full-Text Search engine as a service .... 65
Shutting down the Enhanced Full-Text Search engine ............. 66
Modifying the configuration parameters ................................. 67
Modifying configuration values ............................................... 69
## Contents

Top-level topics ................................................................. 94
Subtopics ............................................................................ 94
Evidence topics ................................................................. 95
Topic and subtopic relationships .......................................... 95
Maximum number of topics .................................................. 96
Topic naming issues ............................................................. 96
Verity query language .......................................................... 97
Query language summary ..................................................... 97
Operator precedence rules ................................................... 101
Sample topic outlines .......................................................... 102
Operator reference .............................................................. 103
   ACCRUE operator .......................................................... 104
   ALL operator .................................................................... 104
   AND operator .................................................................... 104
   ANY operator ..................................................................... 104
   CONTAINS operator ........................................................ 104
   ENDS operator .................................................................. 105
   = (EQUALS) operator ....................................................... 105
   FILTER operator ............................................................. 106
   > (GREATER THAN) operator ............................................ 106
   >= (GREATER THAN OR EQUAL TO) operator ................... 106
   < (LESS THAN) operator .................................................. 106
   <= (LESS THAN OR EQUAL TO) operator ......................... 107
   IN operator ....................................................................... 107
   MATCHES operator ......................................................... 107
   NEAR operator ................................................................... 108
   NEAR/N operator ............................................................. 108
   OR operator ...................................................................... 109
   PARAGRAPH operator ...................................................... 109
   PHRASE operator ............................................................ 109
   SENTENCE operator ........................................................ 109
   SOUNDEX operator .......................................................... 110
   STARTS operator ............................................................. 110
   STEM operator .................................................................. 110
   SUBSTRING operator ....................................................... 110
   THESAURUS operator ....................................................... 111
   TYPO/N operator ............................................................ 111
   WILDCARD operator ........................................................ 111
   Using wildcard special characters .................................... 111
   Searching for nonalphanumeric characters ....................... 112
   WORD operator ............................................................... 113
   Modifier reference ............................................................ 113
   CASE modifier .................................................................. 114
   MANY modifier ............................................................... 114
About This Book

This Users Guide explains how to use the Enhanced Full-Text Search Specialty Data Store product with Sybase® Adaptive Server® Enterprise. The features and functionality of the enhanced version of this product are described in this book.

Audience

This book is for System Administrators who are configuring Adaptive Server for an Enhanced Full-Text Search Specialty Data Store and for users who are performing full-text searches on Adaptive Server data.

How to Use This Book

This book includes these chapters:

• Chapter 1, “Introduction,” provides an overview of the Enhanced Full-Text Search Specialty Data Store.

• Chapter 2, “Understanding the Enhanced Full-Text Search Engine,” describes the components of the Enhanced Full-Text Search Specialty Data Store and how it works.

• Chapter 3, “Configuring Adaptive Server for Full-Text Searches,” describes how to configure Adaptive Server so that Enhanced Full-Text Search Specialty Data Store can perform full-text searches on the databases.

• Chapter 4, “Setting Up Verity Functions,” describes the setup you need to do before you can issue full-text search queries.

• Chapter 5, “Writing Full-Text Search Queries,” describes the components you use to write full-text search queries.

• Chapter 6, “System Administration,” provides information about system administration issues.

• Chapter 7, “Performance and Tuning,” provides information about performance and tuning issues.

• Chapter 8, “Verity Topics,” provides information about configuring the Verity engine.

• Appendix B, “Sample Files,” contains the text of the textsvr.cfg file, describes the sample files included with Enhanced Full-Text Search Specialty Data Store, and discusses issues regarding the sample_text_main.sql script.

• Appendix C, “Unicode Support,” describes how to configure Enhanced Full Text Search Specialty Data Store to use Unicode.

• Appendix D, “Working with XML data,” provides the correct format for XML data indexed into text indexes using fields and zones, and samples of using such data.

Related documents  
The Sybase® Adaptive Server® Enterprise documentation set consists of the following:

• The release bulletin for your platform – contains last-minute information that was too late to be included in the books. A more recent version of the release bulletin may be available on the World Wide Web. To check for critical product or document information that was added after the release of the product CD, use the Sybase Technical Library.

• The Installation Guide for your platform – describes installation, upgrade, and configuration procedures for all Adaptive Server and related Sybase products.

• What’s New in Adaptive Server Enterprise? – describes the new features in Adaptive Server version 15.0, the system changes added to support those features, and changes that may affect your existing applications.

• ASE Replicator Users Guide – describes how to use the Adaptive Server Replicator feature of Adaptive Server to implement basic replication from a primary server to one or more remote Adaptive Servers.

• Component Integration Services Users Guide – explains how to use the Adaptive Server Component Integration Services feature to connect remote Sybase and non-Sybase databases.

• Configuring Adaptive Server Enterprise for your platform – provides instructions for performing specific configuration tasks for Adaptive Server.

• Full-Text Search Specialty Data Store Users Guide – describes how to use the Full-Text Search feature with Verity to search Adaptive Server Enterprise data.
• **Glossary** – defines technical terms used in the Adaptive Server documentation.

• **Historical Server Users Guide** – describes how to use Historical Server to obtain performance information for SQL Server® and Adaptive Server.

• **Java in Adaptive Server Enterprise** – describes how to install and use Java classes as data types, functions, and stored procedures in the Adaptive Server database.

• **Job Scheduler User's Guide** – provides instructions on how to install and configure, and create and schedule jobs on a local or remote Adaptive Server using the command line or a graphical user interface (GUI).

• **Monitor Client Library Programmer’s Guide** – describes how to write Monitor Client Library applications that access Adaptive Server performance data.

• **Monitor Server Users Guide** – describes how to use Monitor Server to obtain performance statistics from SQL Server and Adaptive Server.

• **Performance and Tuning Guide** – is a series of four books that explains how to tune Adaptive Server for maximum performance:
  • **Basics** – the basics for understanding and investigating performance questions in Adaptive Server.
  • **Locking** – describes how the various locking schemas can be used for improving performance in Adaptive Server.
  • **Optimizer and Abstract Plans** – describes how the optimizer processes queries and how abstract plans can be used to change some of the optimizer plans.
  • **Monitoring and Analyzing** – explains how statistics are obtained and used for monitoring and optimizing performance.

• **Quick Reference Guide** – provides a comprehensive listing of the names and syntax for commands, functions, system procedures, extended system procedures, datatypes, and utilities in a pocket-sized book.

• **Reference Manual** – is a series of four books that contains the following detailed Transact-SQL® information:
  • **Building Blocks** – Transact-SQL datatypes, functions, global variables, expressions, identifiers and wildcards, and reserved words.
  • **Commands** – Transact-SQL commands.
- **Procedures** – Transact-SQL system procedures, catalog stored procedures, system extended stored procedures, and dbcc stored procedures.
- **Tables** – Transact-SQL system tables and dbcc tables.
- **System Administration Guide** – provides in-depth information about administering servers and databases. This manual includes instructions and guidelines for managing physical resources, security, user and system databases, and specifying character conversion, international language, and sort order settings.
- **System Tables Diagram** – illustrates system tables and their entity relationships in a poster format. Available only in print version.
- **Transact-SQL Users Guide** – documents Transact-SQL, Sybase’s enhanced version of the relational database language. This manual serves as a textbook for beginning users of the database management system. This manual also contains descriptions of the pubs2 and pubs3 sample databases.
- **Using Adaptive Server Distributed Transaction Management Features** – explains how to configure, use, and troubleshoot Adaptive Server DTM features in distributed transaction processing environments.
- **Using Sybase Failover in a High Availability System** – provides instructions for using Sybase’s Failover to configure an Adaptive Server as a companion server in a high availability system.
- **Unified Agent and Agent Management Console** – Describes the Unified Agent, which provides runtime services to manage, monitor and control distributed Sybase resources.
- **Utility Guide** – documents the Adaptive Server utility programs, such as isql and bcp, which are executed at the operating system level.
- **Web Services Users Guide** – explains how to configure, use, and troubleshoot Web Services for Adaptive Server.
- **XA Interface Integration Guide for CICS, Encina, and TUXEDO** – provides instructions for using the Sybase DTM XA interface with X/Open XA transaction managers.
- **XML Services in Adaptive Server Enterprise** – describes the Sybase native XML processor and the Sybase Java-based XML support, introduces XML in the database, and documents the query and mapping functions that comprise XML Services.
About This Book

Other sources of information

Use the Sybase Getting Started CD, the SyBooks CD, and the Sybase Product Manuals Web site to learn more about your product:

• The Getting Started CD contains release bulletins and installation guides in PDF format, and may also contain other documents or updated information not included on the SyBooks CD. It is included with your software. To read or print documents on the Getting Started CD, you need Adobe Acrobat Reader, which you can download at no charge from the Adobe Web site using a link provided on the CD.

• The SyBooks CD contains product manuals and is included with your software. The Eclipse-based SyBooks browser allows you to access the manuals in an easy-to-use, HTML-based format. Some documentation may be provided in PDF format, which you can access through the PDF directory on the SyBooks CD. To read or print the PDF files, you need Adobe Acrobat Reader.

Refer to the SyBooks Installation Guide on the Getting Started CD, or the README.txt file on the SyBooks CD for instructions on installing and starting SyBooks.

• The Sybase Product Manuals Web site is an online version of the SyBooks CD that you can access using a standard Web browser. In addition to product manuals, you will find links to EBFs/Maintenance, Technical Documents, Case Management, Solved Cases, newsgroups, and the Sybase Developer Network.

To access the Sybase Product Manuals Web site, go to Product Manuals at http://www.sybase.com/support/manuals/.

Sybase certifications on the Web

Technical documentation at the Sybase Web site is updated frequently.

❖ Finding the latest information on product certifications

   1 Point your Web browser to Technical Documents at http://www.sybase.com/support/techdocs/.
   2 Select Products from the navigation bar on the left.
   3 Select a product name from the product list and click Go.
   4 Select the Certification Report filter, specify a time frame, and click Go.
   5 Click a Certification Report title to display the report.
Finding the latest information on component certifications
2. Either select the product family and product under Search by Product; or select the platform and product under Search by Platform.
3. Select Search to display the availability and certification report for the selection.

Creating a personalized view of the Sybase Web site (including support pages)
Set up a MySybase profile. MySybase is a free service that allows you to create a personalized view of Sybase Web pages.
2. Click MySybase and create a MySybase profile.

Finding the latest information on EBFs and software maintenance
2. Select EBFs/Maintenance. If prompted, enter your MySybase user name and password.
3. Select a product.
4. Specify a time frame and click Go. A list of EBFs/Maintenance releases is displayed.
   Padlock icons indicate that you do not have download authorization for certain EBFs/Maintenance releases because you are not registered as a Technical Support Contact. If you have not registered, but have valid information provided by your Sybase representative or through your support contract, click Edit Roles to add the “Technical Support Contact” role to your MySybase profile.
5. Click the Info icon to display the EBFs/Maintenance report, or click the product description to download the software.

Conventions
Directory paths For readability, directory paths in this manual are in UNIX format. On Windows NT, substitute $SYBASE with %SYBASE% and replace slashes (/) with backslash (\). For example, replace this user input:
Formatting SQL statements  SQL is a free-form language: there are no rules about the number of words you can put on a line or where you must break a line. However, for readability, all examples and syntax statements in this manual are formatted so that each clause of a statement begins on a new line. Clauses that have more than one part extend to additional lines, which are indented.

SQL syntax conventions  The conventions for syntax statements in this manual are as follows:

<table>
<thead>
<tr>
<th>Key</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>command</td>
<td>Command names, command option names, utility names, utility flags, and other keywords are in <em>bold Courier</em> in syntax statements and in <em>bold Helvetica</em> in paragraph text.</td>
</tr>
<tr>
<td>variable</td>
<td>Variables, or words that stand for values that you fill in, are in <em>italics</em>.</td>
</tr>
<tr>
<td>{   }</td>
<td>Curly braces indicate that you choose at least one of the enclosed options. Do not include braces in your option.</td>
</tr>
<tr>
<td>[   ]</td>
<td>Brackets mean choosing one or more of the enclosed options is optional. Do not include brackets in your option.</td>
</tr>
<tr>
<td>(   )</td>
<td>Parentheses are to be typed as part of the command.</td>
</tr>
<tr>
<td></td>
<td>The vertical bar means you may select only one of the options shown.</td>
</tr>
<tr>
<td></td>
<td>The comma means you may choose as many of the options shown as you like, separating your choices with commas to be typed as part of the command.</td>
</tr>
</tbody>
</table>

- Syntax statements (displaying the syntax and all options for a command) are printed like this:

  ```
  sp_dropdevice [device_name]
  ```

  or, for a command with more options:

  ```
  select column_name
  from table_name
  where search_conditions
  ```
In syntax statements, keywords (commands) are in normal font and identifiers are in lowercase: normal font for keywords, italics for user-supplied words.

- Examples showing the use of Transact-SQL commands are printed like this:
  
  ```sql
  select * from publishers
  ```

- Examples of output from the computer are printed like this:

<table>
<thead>
<tr>
<th>pub_id</th>
<th>pub_name</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>0736</td>
<td>New Age Books</td>
<td>Boston</td>
<td>MA</td>
</tr>
<tr>
<td>0877</td>
<td>Binnet &amp; Hardley</td>
<td>Washington</td>
<td>DC</td>
</tr>
<tr>
<td>1389</td>
<td>Algodata Infosystems</td>
<td>Berkeley</td>
<td>CA</td>
</tr>
</tbody>
</table>

(3 rows affected)

**Case**  In this manual, most of the examples are in lowercase. However, you can disregard case when typing Transact-SQL keywords. For example, `SELECT`, `Select`, and `select` are the same.

Adaptive Server’s sensitivity to the case of database objects, such as table names, depends on the sort order installed on Adaptive Server. You can change case sensitivity for single-byte character sets by reconfiguring the Adaptive Server sort order.

**Obligatory options (you must choose at least one)**

- **Curly Braces and Vertical Bars**: Choose one and only one option.

  ```
  {die_on_your_feet | live_on_your_knees | live_on_your_feet}
  ```

- **Curly Braces and Commas**: Choose one or more options. If you choose more than one, separate your choices with commas.

  ```
  {cash, check, credit}
  ```

**Optional options [you do not have to choose any]**

- **One Item in Square Brackets**: You don’t have to choose it.

  ```
  [anchovies]
  ```

- **Square Brackets and Vertical Bars**: Choose none or only one.

  ```
  [beans | rice | sweet_potatoes]
  ```

- **Square Brackets and Commas**: Choose none, one, or more than one option. If you choose more than one, separate your choices with commas.

  ```
  [extra_cheese, avocados, sour_cream]
  ```
**Ellipsis: Do it again (and again)....** An ellipsis (…) means that you can repeat the last unit as many times as you like. In this syntax statement, `buy` is a required keyword:

```
buy thing = price [cash | check | credit]
[, thing = price [cash | check | credit]]...
```

You must buy at least one thing and give its price. You may choose a method of payment: one of the items enclosed in square brackets. You may also choose to buy additional things: as many of them as you like. For each thing you buy, give its name, its price, and (optionally) a method of payment.

This document is available in an HTML version that is specialized for accessibility. You can navigate the HTML with an adaptive technology such as a screen reader, or view it with a screen enlarger.

This version of the Enhanced Specialty Data Store and the HTML documentation have been tested for compliance with U.S. government Section 508 Accessibility requirements. Documents that comply with Section 508 generally also meet non-U.S. accessibility guidelines, such as the World Wide Web Consortium (W3C) guidelines for Web sites.

The online help for this product is also provided in HTML, which you can navigate using a screen reader.

**Note** You might need to configure your accessibility tool for optimal use. Some screen readers pronounce text based on its case; for example, they pronounce ALL UPPERCASE TEXT as initials, and MixedCase Text as words. You might find it helpful to configure your tool to announce syntax conventions. Consult the documentation for your tool.

For information about how Sybase supports accessibility, see **Sybase Accessibility** at [http://www.sybase.com/accessibility](http://www.sybase.com/accessibility). The Sybase Accessibility site includes links to information on Section 508 and W3C standards.
CHAPTER 1

Introduction

Enhanced Full-Text Search Specialty Data Store (referred to in this book as the Enhanced Full-Text Search engine) is an Open Server™ application built on the Verity technology that is available in the Verity Developer’s Kit. Adaptive Server connects to the Enhanced Full-Text Search engine through Component Integration Services (CIS), allowing queries written in the Verity query language to perform full-text searches on Adaptive Server data.

This book describes the features and functionality of the Enhanced Full-Text Search Specialty Data Store.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capabilities of the Enhanced Full-Text Search Engine</td>
<td>1</td>
</tr>
<tr>
<td>High availability</td>
<td>3</td>
</tr>
</tbody>
</table>

Capabilities of the Enhanced Full-Text Search Engine

The Enhanced Full-Text Search Specialty Data Store product performs powerful, full-text searches on Adaptive Server data. In Adaptive Server, without the Enhanced Full-Text Search engine, you can search text columns only for data that matches what you specify in a `SELECT` statement. For example, if a table contains documents about dog breeds, and you perform a search on the words “Saint Bernard,” the query produces only the rows that include “Saint Bernard” in the text column.

With the Enhanced Full-Text Search engine, you can expand queries on text columns to:

- Rank the results by order of how often a searched item appears in the selected document. For example, you can obtain a list of document titles that reference the words “Saint Bernard” five or more times.
Capabilities of the Enhanced Full-Text Search Engine

- Select documents in which the words you search for appear within $n$ number of words of each other. For example, you can search only for the documents that include the words “Saint Bernard” and “Swiss Alps” and that appear within 10 words of each other.

- Select documents that include all the search elements you specify within a single paragraph or sentence. For example, you can query the documents that include the words “Saint Bernard” in the same paragraph or sentence as the words “Swiss Alps.”

- Select documents that contain one or more synonyms of the word you specify. For example, you can select documents that discuss “dogs,” and it returns documents that contain the words “dogs,” “canine,” “pooch,” “pup,” and so on.

- Create your own custom thesaurus. For example, you can create a custom thesaurus that includes “working dogs,” “St. Bernard,” “large dogs,” and “European Breeds” as synonyms for “Saint Bernard.”

- Create topics that specify the search criteria for a query. For example, you can create a topic that returns documents that include the phrase “Saint Bernard” or “St. Bernard,” followed by documents that include the phrase “working dogs,” “large dogs,” or “European Breeds.”

- Return documents grouped in clusters to give you a sense of the major topics covered in the documents.

- Select a section of relevant text in a document and search for other, similar documents.

- Index many different document types, such as Microsoft Word, and FrameMaker.

- Sort documents using up to 16 sort orders.

- Integrate backup and restore capabilities.

- Change the value of a configuration parameter using a system procedure.

- Optimize indexes for text searches when your server is inactive, to enhance performance.

- Create additional system management reports for viewing setup information.

- Ability to bring databases online automatically for text searches.
High availability

The Enhanced Full-Text Search product supports Sybase Failover. If an Adaptive Server fails, the Enhanced Full-Text Search accepts connections from the companion server. Additionally, if the Adaptive Server has proxy database support enabled, then both the primary and companion servers can use the Enhanced Full-Text Search at the same time.
Capabilities of the Enhanced Full-Text Search Engine
CHAPTER 2

Understanding the Enhanced Full-Text Search Engine

This chapter describes how an Enhanced Full-Text Search engine works.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components of the Enhanced Full-Text Search engine</td>
<td>5</td>
</tr>
<tr>
<td>How a full-text search works</td>
<td>10</td>
</tr>
</tbody>
</table>

Components of the Enhanced Full-Text Search engine

The Enhanced Full-Text Search engine uses the following components to provide full-text search capabilities:

- Source table
- Verity collections (text index)
- Filters for a variety of document types
- text_db database
- Index table
- text_events table

The source table

The source table is a user table maintained by Adaptive Server. It contains one or more columns using the date, time, text, image, char, varchar, datetime, small datetime, bigint, int, smallint, tinyint, unsigned bigint, unsigned int, unsigned smallint, or unitext datatype, which holds the data to be searched. The source table must have an IDENTITY column or primary key, which is used to join the source table with the id column of an index table during text searches.
Components of the Enhanced Full-Text Search engine

The source table can be a local table, which holds the actual data, or it can be a proxy table that is mapped to remote data using CIS.

The Verity collections

The Enhanced Full-Text Search engine uses the Verity collections, which are located in $SYBASE/$SYBASE_FTS/collections. When you create the text indexes, as described in “Creating the text index and index table” on page 20, Verity creates a collection, which is a directory that implements a text index. This collection is queried by the Enhanced Full-Text Search engine. For more information about Verity collections, see the Verity Web site at http://www.verity.com.

Filters

The text index uses a filter to strip out the tags in a document that are not ASCII text. The Enhanced Full-Text Search engine provides filters for a variety of document types (Microsoft Word, PDF, WordPerfect, SGML, and HTML).

The text_db database

During the installation of the Enhanced Full-Text Search engine, a database named text_db is added to Adaptive Server using the installation script installtextserver, as described in “Running the installtextserver script” on page 14. The database does not contain any user data, but contains two support tables: vesaux and vesauxcol. These tables contain the metadata used by the Enhanced Full-Text Search engine to maintain integrity between the Adaptive Server source tables and the Verity collections.

When updating the collections after an insert, update, or delete is made to an indexed column, the Enhanced Full-Text Search engine queries the vesaux and vesauxcol tables. These tables determine which collections contain the modified columns so that all affected collections are updated. The Enhanced Full-Text Search engine also uses these tables when it is brought online, to make sure that all necessary collections exist.

The vesaux table

The columns in the vesaux table are described in Table 2-1.
The \textit{vesauxcol} table

The columns in the \textit{vesauxcol} table are described in Table 2-2.

<table>
<thead>
<tr>
<th>Column name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>ID of the referenced row in the \textit{vesaux} table</td>
</tr>
<tr>
<td>col_name</td>
<td>Name of the column for which you are searching</td>
</tr>
<tr>
<td>col_type</td>
<td>Column type (date, time, text, image, char, varchar, datetime, smalldatetime; with the Enhanced Full-Text Search engine, also int, smallint, and tinyint)</td>
</tr>
</tbody>
</table>

The \textit{index} table

The \textit{index} table provides a means of locating and searching documents stored in the source table. The \textit{index} table is maintained by the Enhanced Full-Text Search engine and has an \textit{id} column that maps to the \textit{IDENTITY} column or primary key of the corresponding source table. The \textit{IDENTITY} or primary key value from the row in the source table is stored with the data in the Verity collections, which allows the source and index tables to be joined. Although the index table is stored and maintained by the Enhanced Full-Text Search engine, it functions as a proxy table to Adaptive Server through Component Integration Services.

The index table contains special columns, called \textit{pseudo columns}, that are used by the Enhanced Full-Text Search engine to determine the parameters of the search and the location of the text data in the source table. Pseudo columns have no associated physical storage—the values of a pseudo column are valid only for the duration of the query and are removed immediately after the query finishes running.
For example, when you use the score pseudo column in a query, to rank each document according to how well the document matches the query, you may have to use a score of 15 to find references to the phrase “small Saint Bernards” in the text database. This phrase does not occur very often, and a low score value broadens the search to include documents that have a small number of occurrences of the search criteria. However, if you are searching for a phrase that is common, like “large Saint Bernards,” you could use a score of 90, which would limit the search to those documents that have many occurrences of the search criteria.

You use the score column and the other pseudo columns, id, index_any, sort_by, summary, and max_docs, to specify the parameters to include in your search. For a description of the pseudo columns, see “Pseudo columns in the index table” on page 44.

The **text_events** table

Each database containing tables for which there is a text index must contain an events table, which logs inserts, updates, and deletes to indexed columns. The name of this table is text_events. It is used to propagate updated data to the Verity collections.

The columns in the text_events table are described in Table 2-3.

<table>
<thead>
<tr>
<th>Column name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>event_id</td>
<td>IDENTITY column</td>
</tr>
<tr>
<td>id</td>
<td>ID of the row that was updated, inserted, or deleted</td>
</tr>
<tr>
<td>tableid</td>
<td>Name of the table that contains the row that was updated, inserted, or deleted</td>
</tr>
<tr>
<td>columnid</td>
<td>Name of the column on which the text index was created</td>
</tr>
<tr>
<td>event_date</td>
<td>Date and time of the update, insert, or delete</td>
</tr>
<tr>
<td>event_type</td>
<td>Type of update (update, insert, or delete)</td>
</tr>
<tr>
<td>event_status</td>
<td>Indicates whether the update, insert, or delete has been propagated to the collections.</td>
</tr>
<tr>
<td>0</td>
<td>Event Unread</td>
</tr>
<tr>
<td>1</td>
<td>Event Read</td>
</tr>
<tr>
<td>2</td>
<td>Event Succeeded</td>
</tr>
<tr>
<td>3</td>
<td>Event Failed</td>
</tr>
<tr>
<td>srvid</td>
<td>Server ID of the Enhanced Full-Text Search engine maintaining the collection</td>
</tr>
</tbody>
</table>
Relationships between the components

The relationships between the Enhanced Full-Text Search engine components are shown in Figure 2-1.

*Figure 2-1: Components of the Enhanced Full-Text Search engine*
How a full-text search works

To perform a full-text search, you enter a select statement that joins the IDENTITY column or primary key from the source table with the id column of the index table, using pseudo columns as needed to define the search. For example, the following query searches for documents in the blurs table of the pubs2 database in which the word “Greek” appears near the word “Gustibus” (the i_blurs table is the index table):

```
select t1.score, t2.copy
from i_blurs t1, blurs t2
where t1.id=t2.id and t1.score > 20
    and t1.max_docs = 10
    and t1.index_any = "<near>(Greek, Gustibus)"
```

Adaptive Server and the Enhanced Full-Text Search engine split the query processing, as follows:

1. The Enhanced Full-Text Search engine processes the query:
   ```
   select t1.score, t1.id
   from i_blurs t1
   where t1.score > 20
       and t1.max_docs = 10
       and t1.index_any = "<near>(Greek, Gustibus)"
   ```
   The `select` statement includes the Verity operator `near` and the pseudo columns `score`, `max_docs`, and `index_any`. The operator and pseudo columns provide the parameters for the search on the Verity collections—they narrow the result set from the entire `copy` column to the 10 documents in which the words “Greek” and “Gustibus” appear closest to each other.

2. Adaptive Server processes the following `select` statement on the result set that is returned by the Enhanced Full-Text Search engine in step 1:
   ```
   select t1.score, t2.copy
   from i_blurs t1, blurs t2
   where t1.id=t2.id
   ```
   This joins the blurs and i_blurs tables (the source table and the index table, respectively) on the IDENTITY column or primary key of the blurs table and the id column of the i_blurs table.

Figure 2-2 describes how Adaptive Server and the Enhanced Full-Text Search engine process the query.
Figure 2-2: Processing a full-text search query

1. Adaptive Server sends the index query to the Enhanced Full-Text Search engine.
2. The Enhanced Full-Text Search engine processes the Verity operators in the query and produces a result set from the collections.
3. The Enhanced Full-Text Search engine returns the result set to Adaptive Server.
4. Adaptive Server processes the select statement on the local table.
5. Adaptive Server displays the results of the query.
How a full-text search works
CHAPTER 3

Configuring Adaptive Server for Full-Text Searches

This chapter describes how to configure Adaptive Server to perform full-text searches.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuring Adaptive Server for an Enhanced Full-Text Search engine</td>
<td>13</td>
</tr>
<tr>
<td>Creating and maintaining text indexes</td>
<td>19</td>
</tr>
</tbody>
</table>

Configuring Adaptive Server for an Enhanced Full-Text Search engine

The Enhanced Full-Text Search engine is a remote server that Adaptive Server connects to through Component Integration Services (CIS). Before you can use the Enhanced Full-Text Search engine, you must configure Adaptive Server for the Enhanced Full-Text Search engine as follows:

- Enable the enable cis, cis rpc handling and full-text search configuration parameters if you have not done so. You need a license to enable full-text search.
- Run the installtextserver script to define one or more Enhanced Full-Text Search engines.
- Run the installmessages script to install messages for the Enhanced Full-Text Search engine’s system procedures.
- Run the installevent script to create the text_events table in each user database that will contain text indexes.
- Name the local server and restart.
Enabling configuration parameters

To connect to the Enhanced Full-Text Search engine, Adaptive Server must be running with the `enable cis` and `cis rpc handling` configuration parameters enabled. If those parameters are not enabled, log in to Adaptive Server using `isql` and use `sp_configure` to enable them. For example:

```sql
exec sp_configure "enable cis", 1
exec sp_configure "cis rpc handling", 1
exec sp_configure "enable full-text search", 1
```

If you have made changes to `enable cis`, messages display stating that you must restart Adaptive Server for the new configuration parameter to take effect.

Running the `installtextserver` script

The `installtextserver` script:

- Defines the Enhanced Full-Text Search engine as a remote server of server class `sds` to Adaptive Server.
- Creates a database for storing text index metadata. For more information about this database, see “The text_db database” on page 6.
- Installs the system procedures required by the Enhanced Full-Text Search engine.

Run the `installtextserver` script only once (see “Starting the `installtextserver` script” on page 16). To add Enhanced Full-Text Search engines at a later time, use `sp_addserver`. See “Configuring multiple Enhanced Full-Text Search engines” on page 84 for more information about `sp_addserver`.

**Note** If you are putting the `installtextserver` script onto the `model` database, you must increase the size of the `model` database to at least 3MB prior to running the script.

All Enhanced Full-Text Search engines use the same database for storing text index metadata. This database is referred to as the `text_db` database, the default name.

For a list and description of the system procedures added with the `installtextserver` script, see Appendix A, “System Procedures.”
CHAPTER 3  Configuring Adaptive Server for Full-Text Searches

Editing the installtextserver script

The installtextserver script is located in the $SYBASE/$SYBASE_FTS/scripts directory. Use a text editor (such as vi or emacs) to open the script, and make your edits. The edits you can make are as follows:

- Changing the name of the text_db database. If you use a different name, replace all occurrences of text_db with the appropriate name.

  **Note** If you change the name of the text_db database, you must also change the name in the defaultDb configuration parameter (see “Modifying the configuration parameters” on page 67).

- Changing the name of the Enhanced Full-Text Search engine. By default, the installtextserver script defines a Enhanced Full-Text Search engine named “textsvr.” If your Enhanced Full-Text Search engine is named differently, edit this script so that it defines the correct server name.

- Adding multiple Enhanced Full-Text Search engines (for information on how this can enhance performance, see “Configuring multiple Enhanced Full-Text Search engines” on page 84). If you are initially defining more than one Enhanced Full-Text Search engine, edit the installtextserver script so that it includes all the Enhanced Full-Text Search engine definitions. installtextserver includes the following section for naming the Enhanced Full-Text Search engine you are configuring (“textsvr” by default):

  ```
  /*
   ** Add the text server
   */
   exec sp_addserver textsvr, sds, textsvr
   go
  ```

  Add an entry for each Enhanced Full-Text Search engine you are configuring. For example, if you are configuring three Enhanced Full-Text Search engines named KRAZYKAT, OFFICAPUP, and MOUSE, replace the default “textsvr” line with the following lines:

  ```
  exec sp_addserver KRAZYKAT, sds, KRAZYKAT
  exec sp_addserver OFFICAPUP, sds, OFFICAPUP
  exec sp_addserver MOUSE, sds, MOUSE
  go
  ```
If you use OmniConnect to communicate with the Enhanced Full-Text Search engine, change the server name specification in the sp_addobjectdef calls for the vesaux and vesauxcol tables to a valid remote server. For example, if your remote server is named REMOTE, change the lines:

```sql
exec sp_addobjectdef
"vesaux","SYBASE.master.dbo.vesaux","table"
exec sp_addobjectdef
"vesauxcol","SYBASE.master.dbo.vesauxcol","table"
```
to:

```sql
exec sp_addobjectdef
"vesaux","REMOTE.master.dbo.vesaux","table"
exec sp_addobjectdef
"vesauxcol","REMOTE.master.dbo.vesauxcol","table"
```

**Starting the `installtextserver` script**

Use `isql` to run the `installtextserver` script. For example, to run the `installtextserver` script in an Adaptive Server named MYSVR, enter:

```bash
isql -Usa -P -SMYSVR -i
$SYBASE/$SYBASE_FTS/scripts/installtextserver
```

**Running the `installmessages` script**

The Enhanced Full-Text Search engine has its own set of system procedure messages that you must install in Adaptive Server. Use the `installmessages` script to install the messages. You run the `installmessages` script only once, even if you have multiple Enhanced Full-Text Search engines.

For example, to run the `installmessages` script in a server named MYSVR, enter:

```bash
isql -Usa -P -SMYSVR -i
$SYBASE/$SYBASE_FTS/scripts/installmessages
```
Running the *installevent* script

Each database containing tables referenced by a text index must contain a `text_events` table, which logs inserts, updates, and deletes to indexed columns. It is used to propagate updated data to the Verity collections.

Run the `installevent` script, as described below, to create the `text_events` table and associated system procedures in a database. Use the `installevent` script as follows:

- If all databases require text indexes, run the `installevent` script to create a `text_events` table in the `model` database. Each newly created database will then have a `text_events` table. To add a `text_events` table to existing databases, edit the script as described below to create the `text_events` table in the existing user database.

- If not all databases have text indexes, use the `installevent` script as a sample. For each existing database and each new database that includes tables that require text indexing, run the `installevent` script. You must edit the script as described below, to create the `text_events` table in the correct user database.

**Note** If a `text_events` table does not exist in a database that includes source tables that require text indexing, changes to the source table are not propagated to the Verity collections.

Editing the *installevent* script

The `installevent` script is located in the `$SYBASE/SSYBASE_FTS/scripts` directory. Use a text editor (such as `vi` or `emacs`) to open the script, and make the edits. The edits you can make are:

- Changing the user database name. The `installevent` script creates an events table (named `text_events`) and associated system procedures in the `model` database. The `model` database is the default database. To install the `text_events` table in an existing user database, edit the script and replace all references to `model` with the user database name.
Configuring Adaptive Server for an Enhanced Full-Text Search engine

- Changing the text_db database name. If your database for storing text index metadata is named something other than text_db, replace all references to text_db with the appropriate name.

  **Note** The name of the text_db database must be the same as the name in the defaultDb configuration parameter (see “Modifying the configuration parameters” on page 67).

Running the `installevent` script

Using isql, run the installevent script to install the text_events table and related system procedures in Adaptive Server. For example, to run the installevent script in a server named MYSVR, enter:

```
isql -Usa -P -SMYSVR -i
$SYBASE/$SYBASE_FTS/scripts/installevent
```

  **Note** The text_db database must exist before you run the installevent script. If it does not exist, run the installtextserver script first.

  **Note** Before installing the EFTS-12_5 scripts on 64-bit platforms (installtextserver, installevents, installmessages) increase the size of the tempdb and the model databases from the default size to 3MB each.

Naming the local Adaptive Server

When using the Enhanced Full-Text Search engine with Adaptive Server 12.5 and later, you must name the local Adaptive Server using `sp_addserver <servername>, local`. After issuing `sp_addserver`, you must restart the local Adaptive Server. Do not install any system stored procedures in the model database. They should be installed in sybsystemprocs.
Creating and maintaining text indexes

Before the Enhanced Full-Text Search engine can process full-text searches, you must create text indexes for the source tables in the user database. After the text indexes are created, you must update them when the source data changes to keep the text indexes current. To create and maintain the text indexes:

1. Set up the source table for indexing (see “Setting up source tables for indexing” on page 19).
2. Create the text indexes and index tables (see “Creating the text index and index table” on page 20).
3. Bring the databases online for full-text searches (see “Bringing the database online for full-text searches” on page 22).
4. Propagate changes in the user data to the text indexes (see “Propagating changes to the text index” on page 23).
5. If you are replicating text indexes, set up text indexing in the destination database (see “Replicating text indexes” on page 24).

For an example of setting up a text index, see the sample script `sample_text_main.sql` in the `$SYBASE/$SYBASE_FTS/sample/scripts` directory.

Setting up source tables for indexing

The source table contains the data on which you perform searches (for example, the `blurbs` table in the `pubs2` database). For more information, see “The source table” on page 5.

Before you can create text indexes on a source table, you must:

- Verify that the source table has an IDENTITY column or primary key. If not, alter the table and add an IDENTITY column.
- Create a unique index on the primary key or IDENTITY column (optional).
Creating and maintaining text indexes

Every source table must contain an IDENTITY column or primary key to uniquely identify each row and provide a means of joining the index table and the source table. When you create a text index, the IDENTITY column or primary key is passed with the indexed columns to the Enhanced Full-Text Search engine. The IDENTITY column or primary key value is stored in the text index and is mapped to the id column in the index table. If a table does not have a primary key an IDENTITY column can be added to the table.

Adding an IDENTITY column to a source table

To create an IDENTITY column in a table named composers, define the table as follows:

```sql
create table composers (
    id numeric(m,n) identity,
    comp.fname char(30) not null,
    comp.lname char(30) not null,
    text_col text
)
```

where m ≤ 38 and n always = 0.

To add an IDENTITY column to an existing table, enter:

```sql
alter table table_name add id numeric(10,0) identity
```

Adding a unique index to an IDENTITY column

For optimum performance, Sybase recommends that you create a unique index on the IDENTITY column. For example, to create a unique index named comp_id on the IDENTITY column created above, enter:

```sql
create unique index comp_id
on composers(id)
```

For more information about creating a unique index, see Chapter 11, “Creating Indexes on Tables,” in the Transact-SQL Users Guide.

Creating the text index and index table

Use sp_create_text_index to create text indexes. sp_create_text_index:

- Updates the vesaux and vesauxcol tables in the text_db database
- Creates the text index (Verity collections)
CHAPTER 3 Configuring Adaptive Server for Full-Text Searches

- Populates the Verity collections
- Creates the index table in the user database where the source table is located

**Note** The Enhance Full-Text Search engine must be running to successfully execute `sp_create_text_index`. For information on starting and stopping the Enhanced Full-Text Search engine, see Chapter 6, “System Administration.”

The text index can contain up to 16 columns. Columns of the following datatypes can be indexed: char, varchar, nchar, nvarchar, date, time, timestamp, text, image, datetime, smalldatetime, int, smallint, tinyint, unichar, and univarchar.

For example, to create a text index and an index table named `i_blurbs` for the `copy` column in the `blurbs` table in `pubs2` on KRAZYKAT, enter:

```
sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", "", "copy"
```

where:
- KRAZYKAT is the name of the Enhanced Full-Text Search engine.
- i_blurbs is the name of the index table and text index you are creating.
- blurbs is the source table on which you are creating the text indexes.
- " " is a placeholder for text index creation options.
- copy is the column in the blurbs table that you are indexing.

See `sp_create_text_index` on page 143 for more information.

**Note** Make sure the text_db database name in the configuration file (listed after the `defaultDb` parameter) matches the database name in Adaptive Server. If they do not match, the text index cannot be created. Also, verify that the text_events table exists in the user database. If it does not exist, run the `installevent` script for that database (see “Running the installevrent script” on page 17).

Populating the Verity collections can take a few minutes or several hours, depending on the amount of data you are indexing. You may want to perform this step when the server is not being heavily used. Increasing the `batch_size` configuration parameter also expedites the process. See “batch_size” on page 82 for more information.

**Note** Do not rename an index; the Verity collection will not be renamed.
Creating and maintaining text indexes

Specifying multiple columns when creating a text index

When you create a text index on two or more columns, each column in the text index is placed into its own document zone. The name of the zone is the name of the column. For example, to create a text index and an index table named i_blurbs for both the copy column and the au_id column in the blurbs table in pubs2 on KRAZYKAT, enter:

```
sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", "", "copy", "au_id"
```

This creates two zones in the text index named “copy” and “au_id.” When you issue a query against the i_blurbs text index, the search includes the copy and au_id columns. However, you can limit your search to a particular column by using the in operator to specify a document zone (for more information, see “in” on page 53).

Granting permissions on text index proxy tables

After a text index is created via `sp_create_text_index`, the administrator must explicitly grant select permissions on the text index proxy table to users or groups who will query the text index.

The user must have select permission on all columns in the text index source in order to be granted select permission on the text index. If the user does not have select permission on even one of the columns in the text index source, the administrator cannot grant select permission on the text index.

If any of the columns in the text index source table are encrypted columns, the administrator should grant select permissions on the text index only to users who have decrypt permission on the encrypted columns. If the user has select permissions on the encrypted columns, but nodecrypt permissions, the administrator should not grant select permission to the text index.

Bringing the database online for full-text searches

With the Enhanced Full-Text Search engine, the database is automatically brought online when the auto_online configuration parameter is set to 1.

When you bring a database online, the Enhanced Full-Text Search engine initializes the internal Verity structures and confirms that the Verity collections exist.
Use `sp_text_online` to bring a database online for full-text searches if it is not automatically brought online. For example, to bring the `pubs2` database online before issuing full-text searches on the `blurbs` table in a Enhanced Full-Text Search engine named KRAZYKAT, enter:

```
sp_text_online KRAZYKAT, pubs2
```

This message appears:

```
Database 'pubs2' is now online
```

The `pubs2` database is now available for performing full-text searches.

See `sp_text_online` on page 162 for more information.

**Propagating changes to the text index**

When you insert, update, or delete data in your source table, the text indexes are not updated automatically. After you update data, run `sp_refresh_text_index` to log the changes to the `text_events` table. Then, run `sp_text_notify` to notify the Enhanced Full-Text Search engine that changes need to be processed. The Enhanced Full-Text Search engine then connects to Adaptive Server, reads the entries in the `text_events` table, determines which indexes, tables, and rows are affected, and updates the appropriate collections.

See `sp_refresh_text_index` on page 149 and `sp_text_notify` on page 161 for more information on these system procedures.

To have `sp_refresh_text_index` run automatically after each insert, update, or delete, you can create triggers on your source tables, as follows:

- Create a trigger that runs `sp_refresh_text_index` after a delete operation.
- Create a trigger that runs `sp_refresh_text_index` after an insert operation.
- Create a trigger that runs `sp_refresh_text_index` after an update operation to an indexed column.

Triggers are not fired when you use `writetext` to update a text column. To have `sp_refresh_text_index` automatically run after a `writetext`:

- Set up a non-text column and update that column after each `writetext`.
- Create a trigger on the non-text column to run `sp_refresh_text_index`. Since the Enhanced Full-Text Search engine reinserts the entire row when you issue `sp_text_notify`, the update to the text column gets propagated to the text index.
For examples of each of these triggers, see the sample script sample_text_main.sql in the $SYBASE/$SYBASE_FTS/sample/scripts directory.

### Replicating text indexes

To replicate tables that have text indexes, follow these guidelines:

- Create the table definition in the destination database.
- Run the `installevent` script to create the `text_events` table in the destination database, if the `text_events` table does not already exist (see “Running the `installevent` script” on page 17).
- Run `sp_create_text_index` to create the text index on the empty table in the destination database (see “Creating the text index and index table” on page 20).
- Create triggers for running `sp_create_text_index` to insert entries into the `text_events` table whenever you insert, update, or delete data into the table (see “Propagating changes to the text index” on page 23).
- Create the replication definition in the Replication Server. This replicates all the data in the source table to the destination table. See the [Replication Server Administration Guide](#) for more details.
- Run `sp_text_notify` to update the text index; run `sp_text_notify` periodically to process changes to the destination table (see "Propagating changes to the text index" on page 23).

**Note** You must issue an update against a non-text column whenever a `writetext` command is performed. This ensures that the trigger that inserts data into the `text_events` table is fired.

### Example: enabling a new database for text searches

This example describes the steps for creating a text index on the `plot` column of the `reviews` table in the `movies` database. This process assumes that:

- You have created a `reviews` table in a new database named `movies` on the MYSVR server
- The `reviews` table has a column named `plot` that you are going to index
• Adaptive Server and the Enhanced Full-Text Search engine named MYTXTSVR have been configured to connect to each other

Step 1. Verifying that the text_events table exists

Each database containing tables referenced by a text index must contain a text_events table, which logs inserts, updates, and deletes to indexed columns.

If a text_events table is in your model database, it will be in all new databases. If a text_events table is not in your model database, run the installevent script to install the text_events table in the new database. For example, to install the text_events table in the movies database:

• Save the installevent script as installeventmovies.
• Edit the script to replace all references to the word model with the word movies.
• Run the script as follows:

```sql
isql -Usa -P -SMYSVR -i $SYBASE/$SYBASE_FTS/scripts/installeventmovies
```

See “Running the installevent script” on page 17 for information on installing the text_events table.

Step 2. Checking for an IDENTITY column or primary key

Every source table must contain an IDENTITY column or primary key, which uniquely identifies each row and provides a means of joining the index table and the source table.

For example, to add an IDENTITY column to the reviews table, enter:

```sql
alter table reviews add id numeric(10,0) identity
```

See “Adding an IDENTITY column to a source table” on page 20 for more information.

Step 3. Creating a unique index on the IDENTITY column

This step is optional. To enhance performance, Sybase recommends that you create a unique index that contains only the IDENTITY column. For example, to create a unique index named reviews_id on the IDENTITY column created in the previous procedure, issue:

```sql
create unique index reviews_id on reviews(id)
```
Creating and maintaining text indexes

For more information about creating a unique index, see Chapter 11, “Creating Indexes on Tables,” of the Transact-SQL Users Guide.

Step 4. Creating the text index and index table

The source tables in the user database must be indexed so that you can perform full-text searches. For example, to create a text index and an index table named reviews_idx for the plot column in the reviews table, enter:

\[
\text{sp\_create\_text\_index } \text{"MYTXTSVR"}, \text{"reviews\_idx"}, \\
\text{"reviews"}, \text{""}, \text{"plot"}
\]

The reviews table is now available for running full-text searches.
See sp_create_text_index on page 143 for more information.

Step 5. Bringing the database online for a full-text search

To bring the movies database online for the Enhanced Full-Text Search engine named MYTXTSVR, enter:

\[
\text{sp\_text\_online } \text{MYTXTSVR}, \text{movies}
\]

**Note** Omit this step if you have auto_online set to “1”.

See sp_text_online on page 162 for more information.

Indexing the euro symbol

The euro symbol can be indexed and returned properly if the following configuration guidelines are followed. Adaptive Server must have the utf8 charset installed. Enhanced Full-Text Search must have the vdkLanguage set to <language>x and the vdkCharset left blank. For example:

\[
\text{ASE 12.5.x charset = utf8} \\
\text{EFTS 12.5.x vdkLanguage = englishx} \\
\text{EFTS 12.5.x vdkCharset =}
\]
CHAPTER 4

Setting Up Verity Functions

This chapter describes the setup required before you can write queries with certain Verity functionality.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enabling query-by-example, summarization, and clustering</td>
<td>27</td>
</tr>
<tr>
<td>Setting up a column to use as a sort specification</td>
<td>30</td>
</tr>
<tr>
<td>Using filters on text that contains tags</td>
<td>32</td>
</tr>
<tr>
<td>Creating a custom thesaurus</td>
<td>34</td>
</tr>
<tr>
<td>Creating topics</td>
<td>38</td>
</tr>
</tbody>
</table>

Enabling query-by-example, summarization, and clustering

The *style.prm* file specifies additional data to include in the text indexes to support the following functionality:

- Query-by-example – retrieves documents that are similar to a phrase (see “like” on page 54 for more information).

**Note** The text indexes need additional data to support phrases in the query-by-example specification of the `like` operator. If you use a document in the query-by-example specification, additional data is not required.

- Summarization – returns summaries of documents rather than entire documents (see “Using the summary column to summarize documents” on page 47 for more information).

- Clustering – groups documents in result sets by subtopic. See “Using pseudo columns to request clustered result sets” on page 48 for more information.
Enabling query-by-example, summarization, and clustering

You can enable these features for all text indexes by editing the master style.prm file, or you can enable them for an individual text index by editing its style.prm file. Both methods are described below.

Query-by-example and clustering

To use phrases in a query-by-example specification and to use clustering, you must enable the generation of document feature vectors at indexing time. To do this, uncomment the following line in the style.prm file:

```
$define DOC-FEATURES "TF"
```

Summarization

To configure the Enhanced Full-Text Search engine for summarization, uncomment one of the following lines that starts with “#$define” in the style.prm file:

```
#$define DOC-SUMMARIES   "XS MaxSents 3 MaxBytes 255"
#$define DOC-SUMMARIES   "LS MaxSents 4 MaxBytes 255"
#$define DOC-SUMMARIES   "LB MaxBytes 150"
```

Each of those lines reflects a different level of summarization. You can specify how many bytes of data you want the Enhanced Full-Text Search engine to display, by altering the numbers at the ends of these lines. For example, if you want only the first 233 bytes of data summarized, edit the script to read:

```
$define DOC-SUMMARIES   "LS MaxSents 4 MaxBytes 233"
```

The maximum number of bytes displayed is 255. Any number greater than that is truncated to 255.

Editing the master style.prm file

Note The location of the master style.prm file was moved starting with the EFTS 12.5.2 release. Edits made to the style.prm file located in $SYBASE/$SYBASE_FTS/verity/common/style will be ignored. The new location is specified below.
The master style.prm file is located in $SYBASE/$SYBASE_FTS/verity/common/styles/txtsvr. It contains the default Enhanced Full-Text Search engine style parameters. Edit this file to configure the Enhanced Full-Text Search engine so that all tables on which you create text indexes allow clustering and literal text in your query-by-example specifications, or summarization. Uncomment the applicable lines as described above.

Note If you have existing text indexes, you must re-create the text index with these features enabled as described in “Editing individual style.prm files” on page 29.

Editing individual style.prm files

Perform the following steps to configure the Enhanced Full-Text Search engine so that the individual text index allows clustering and literal text in your query-by-example specifications, or summarization:

1 Create the text index using sp_create_text_index. Use the word “empty” in the option_string parameter so that the style.prm file is created for the text index, but the Verity collections are not populated with data. For example, if you are enabling clustering for the copy column of the blurbs table, use the following syntax:

   sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", "empty", "copy"

   Note If the text index already exists, omit this step. You do not need to create the text index again.

2 Use sp_drop_text_index to drop the text index associated with the style.prm file you are editing.

   For example, to drop the text index created in step 1, enter:

   sp_drop_text_index "blurbs.i_blurbs"

3 Edit the style.prm file that exists for the text index. The style.prm file for an existing collection is located in $SYBASE/$SYBASE_FTS/collections/db.owner.index/style.
Setting up a column to use as a sort specification

For example, if you create a text index called i_blurbs on the pubs2 database, the full path to these files is the following, where db.owner.index is the database, the database owner, and the index created with sp_create_text_index:

$SYBASE/$SYBASE_FTS/collections/pubs2.dbo.i_blurbs/style

4 Uncomment the applicable lines as described above. For example, to enable clustering, uncomment the following line:

$define DOC-FEATURES "TF"

5 Re-create the text index you dropped in step 2. For example, to re-create the i_blurbs text index, enter:

sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", ",", "copy"

Setting up a column to use as a sort specification

Before you can sort by specific columns, you must modify the style.vgw and style.ufl files. (For information on including a column in a sort specification, see “Using the sort_by column to specify a sort order” on page 46.) Both files are in the following, where db.owner.index is the database, the database owner, and the index created using sp_create_text_index:

$SYBASE/$SYBASE_FTS/collections/db.owner.index/style

For example, if you created a text index called i_blurbs on the pubs2 database, the full path to those files would be

$SYBASE/$SYBASE_FTS/collections/pubs2.dbo.i_blurbs/style

To edit the style.vgw and style.ufl files, follow these steps:

1 Drop the text index that contains the columns for which you are adding definitions. Dropping the text index does not drop the collection directory.

For example, to add definitions for the copy column in the blurbs table, use the following command to drop the text index:

sp_drop_text_index i_blurbs

2 Edit the style.vgw file. Following this line:

dda "SybaseTextServer"

Add an entry for the column you are defining. The syntax is as follows, where column_number is the number of the column you are defining:
Column numbers start with 0; if you want the first column to be sorted, specify “f0”; to sort the second column, specify “f1”; to sort the third column, specify “f2”, and so on.

For example, to define the first column in a table, the syntax is:

```plaintext
table: DOCUMENTS
{
    copy: f0 copy_f0
}
```

Then, your `style.vgw` file will be similar to this:

```plaintext
# # Sybase Text Server Gateway
# $control: 1
gateway:
{
    dda: "SybaseTextServer"
    {
        copy: f0 copy_f0
    }
}
```

3 Edit the `style.ufl` file by adding the column definition for a data table named `fts`. The syntax is:

```plaintext
data-table: fts
{
    fixwidth: copy_f0 column_number precision datatype
}
```

Column numbers start with 0; if you want the first column to be sorted, specify “f0”; to sort the second column, specify “f1”; to sort the third column, specify “f2”, and so on. For example, to add a definition for the first column of a table, with a precision of 4, and a datatype of `date`, enter:

```plaintext
data-table: fts
{
    fixwidth: copy_f0 4 date
}
```

Similarly, to add a definition for the second column of a table with a precision of 10, and a datatype of `character`, enter:
Using filters on text that contains tags

To perform accurate searches on documents that contain tags (such as HTML or Post Script), the text index must use a filter to strip out the tags. The Enhanced Full-Text Search engine provides filters for a variety of document types (Microsoft Word, FrameMaker, WordPerfect, SGML, HTML, and others).

When you create the text index to use a filter, the data for each type of tag in the document is placed into its own document zone. For example, if you have a tag called “chapter,” all chapter names are placed into one document zone. You can issue a query that searches the entire document, or that searches only for data in the “chapter” zone (for more information, see “in” on page 53).

To create a text index that uses a filter, modify the style.dft file for that text index:

1. Create the text index using `sp_create_text_index`. Use the word “empty” in the `option_string` parameter so that the style.dft file is created for the text index, but the Verity collections are not populated with data. For example, to create a text index for the `copy` column of the `blurbs` table, use the following syntax:

   ```
   sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", "empty", "copy"
   ```

2. Drop the text index that you create in step 1. This drops the text index, but not the `style.dft` file. For example, use the following command to drop the `i_blurbs` text index:

   ```
   sp_drop_text_index i_blurbs
   ```

3. Edit the `style.dft` file. The `style.dft` file is in the directory `$SYBASE/$SYBASE_FTS/collections/db.owner.index/style`, where `db.owner.index` is the database, the database owner, and the index created using `sp_create_text_index`. For example, if you created a text index called `i_blurbs` on the `pubs2` database, the full path to the `style.dft` file would be `$SYBASE/$SYBASE_FTS/collections/pubs2.dbo.i_blurbs/style`
Following this line:

```
field: f0
```

add syntax to use a filter:

- For SGML documents, use:
  
  ```
  /filter="zone -nocharmap"
  ```

- For HTML documents, use:
  
  ```
  /filter="zone -html -nocharmap"
  ```

Use the following syntax for all document types:

```
/filter="universal"
```

For example, your `style.dft` file for an SGML document will look like this:

```
$control: 1
dft: 
{
   field: f0
   /filter="zone -nocharmap"
   field: f1
   field: f2
   .
   .
   field: f15
}
```

Your `style.dft` file for an SGML document will look like this:

```
$control: 1

dft:
{
   field: f0
   /filter="universal"
   field: f1
   field: f2
   .
   .
   field: f15
```
Creating a custom thesaurus

{ 

**Note** Use getsend to load the database with document data. getsend takes the following arguments: database, table, column and row id. Insert a null value for the rowid for each row of text you want to insert. getsend must insert into an image column for filtering to work. For more information on getsend, refer to the README.TXT file and getsend.c file in $SYBASE/$SYBASE_FTS/sample/source directory.

4 Re-create the index, using `sp_create_text_index`. For example:

```
sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", ",", "copy"
```

Creating a custom thesaurus

The Verity thesaurus operator expands a search to include the specified word and its synonyms (for information on using the thesaurus operator, see “thesaurus” on page 57). You can create a custom thesaurus that contains application-specific synonyms to use in place of the default thesaurus.

For example, the default English language thesaurus contains these words as synonyms for “money:” “cash,” “currency,” “lucre,” “wampum,” and “greenbacks.” You can create a custom thesaurus that contains a different set of synonyms for “money,” such as: “bid,” “tokens,” “credit,” “asset,” and “verbal offer.”

To create a custom thesaurus:

1 Make a list of the synonyms that you will use with your application. It may help to examine the default thesaurus (see “Examining the default thesaurus (optional)” on page 35).

2 Create a control file that contains the synonyms you are defining for your custom thesaurus (see “Creating the control file” on page 35).

3 Create the custom thesaurus using the mksyd utility (see “Creating the thesaurus” on page 37). The mksyd utility is located in $SYBASE/$SYBASE_FTS/verity/<verity_platform_directory>/bin.

   This uses the control file as input.

4 Replace the default thesaurus with your custom thesaurus (see “Replacing the default thesaurus with the custom thesaurus” on page 37).
For more information on “Custom Thesaurus Support” and the mksyd utility, see the Verity Web site at http://www.verity.com.

Two sample files illustrate how to set up and use a custom thesaurus:

- `sample_text_thesaurus.ctl` is a sample control file.
- `sample_text_thesaurus.sql` issues queries against the custom thesaurus defined in the sample control file.

These files are in the `$/SYBASE/$SYBASE_FTS/sample/scripts` directory.

### Examining the default thesaurus (optional)

A control file contains all the synonym definitions for a thesaurus. To examine the default thesaurus, create its control file using the mksyd utility. The mksyd utility is located in

`$/SYBASE/SYBASE_FTS/verity/<verity_platform_directory>/bin`.

Use the syntax:

```
mksyd -dump -syd
  $SYBASE/$SYBASE_FTS/verity/common/vdkLanguage/vdk20.syd
    -f
      work_location/control_file.ctl
```

Where:

- `vdkLanguage` – is the value of the vdkLanguage configuration parameter (for example, “english”).
- `work_location` – is the directory where you want to place the control file.
- `control_file` – is the name of the control file you are creating from the default thesaurus.

To view the default synonym lists, examine the control file (`control_file.ctl`) that is created.

### Creating the control file

Create a control file that contains the new synonyms for your custom thesaurus. The control file is an ASCII text file in a structured format. Using a text editor (such as vi or emacs), either:
Creating a custom thesaurus

- Edit the control file from the default thesaurus and add new synonyms to the existing thesaurus (see “Examining the default thesaurus (optional)” on page 35), or
- Create a new control file that includes only your synonyms.

Control file syntax

The control file contains synonym list definitions in a synonyms: statement. For example, the following is a control file named colors.ctl:

```plaintext
$control: 1
synonyms:
  {list: "red, ruby, scarlet, fuchsia,\ magneta"
   list: "electric blue <or> azure"
   /keys = "lapis"
  }
$$
```

The synonyms: statement includes:

- list: keywords that specify the start of a synonym list. The synonyms in the list are either in query form or in a list of words or phrases separated by commas.

- Each list can optionally have a /keys modifier that specifies one or more keys separated by commas. In the example above, no keys are specified in the first “list.” This means the list is found when the thesaurus is queried for the word “red,” “ruby,” “scarlet,” “fuchsia,” or “magenta.” The second “list” uses the /keys modifier to specify one key. This means the words or phrases in the list satisfy a query only when you specify <thesaurus>lapis.</thesaurus>

**Note** If you use emacs to build a synonym list and any of your lists are longer than one line, turn off auto-fill mode. If you separate your list into multiple lines, include a backslash (\) at the end of each line so that the lines are treated as one list.

For more complex examples of control files, see the Verity Web site.
Creating the thesaurus

The mksyd utility creates the custom thesaurus using a control file as input. It is located in:

$SYBASE/$SYBASE_FTS/verity/<verity_platform_directory>/bin

Run, or define an alias to run, mksyd from this bin directory. Create your custom thesaurus in any work directory.

The mksyd syntax for creating a custom thesaurus is:

mksyd -f control_file.ctl -syd custom_thesaurus.syd

Where:

• control_file – is the name of the control file you created in the previous section.

• custom_thesaurus – is the name of the custom thesaurus you are creating.

For example, to execute the mksyd utility reading the sample control file defined above, and directing output to a work directory, use:

mksyd -f /usr/u/sybase/thesaurus/colors.ctl
-syd /usr/u/sybase/thesaurus/custom.syd

Replacing the default thesaurus with the custom thesaurus

The default thesaurus named vdk20.syd is located in

$SYBASE/$SYBASE_FTS/verity/common/vdkLanguage, where vdkLanguage is the value of the vdkLanguage configuration parameter (for example, the English directory is $SYBASE/$SYBASE_FTS/verity/common/english). Each application and user reading from this location at runtime uses this thesaurus.

To replace it with your custom thesaurus:

1 Back up the default thesaurus before replacing it with the custom thesaurus. For example:

mv /$SYBASE/$SYBASE_FTS/verity/common/english/vdk20.syd default.syd

2 Replace the vdk20.syd file with your custom thesaurus. For example:

cp custom.syd /$SYBASE/$SYBASE_FTS/verity/common/english/vdk20.syd

3 Restart your Examine the control file (control_file.ctl); no configuration file changes are required. The thesaurus is read from this location when the Examine the control file (control_file.ctl) is started, not when a query is executed.
Queries using the thesaurus operator will now use the custom thesaurus.

Creating topics

This section provides a condensed overview of Verity topics. Topics are discussed in detail in Chapter 8, “Verity Topics.”

A topic is a grouping of information related to a concept or subject area. With topic definitions in place, a user can perform searches on the topic instead of having to write queries with complex syntax.

The user creates topics, which can be combinations of words and phrases, Verity operators and modifiers, and weight values. Then, any user can query the topic.

Before you create topics, determine your application requirements, and establish standards for naming conventions and for the location of the following:

- Outline files – contains the topic definitions. Each topic has its own outline file.
- Topic set directories – contains the compiled topic. Each topic has its own topic set directory.
- Knowledge base map file – contains pointers to the topic set directories.

To implement topics, perform the following steps:

1. Create one or more outline input files to define your topics (see “Creating an outline file” on page 39). Each outline file is used to populate one topic set.

2. Create and populate a topic set directory, using the mktopics utility (see “Creating a topic set directory” on page 40). The mktopics utility is located in $SYBASE/$SYBASE_FTS/verity/<verity_platform_directory>/bin.

   Each topic set directory is populated based on one topic outline input file.

3. Create a knowledge base map, specifying the locations of one or more topic set directories (see “Creating a knowledge base map” on page 40).

4. Set the knowledge_base configuration parameter to point to the location of the knowledge base map (see “Defining the location of the knowledge base map” on page 41).
5 Execute queries against defined topics.

The following sample files illustrate the topics feature:

- `sample_text_topics.otl` is a sample outline file.
- `sample_text_topics.kbm` is a sample knowledge base map.
- `sample_text_topics.sql` issues queries using defined topics.

These files are in the `$SYBASE/$SYBASE_FTS/sample/scripts` directory.

### Creating an outline file

A topic outline file specifies all the combinations of words and phrases, Verity operators and modifiers, and weight values that the search engine uses when you issue a query using the topic. The outline file is an ASCII text file in a structured format.

For example, the following outline file defines the topic “saint-bernard”:

```
$control: 1
saint-bernard <accrue>
*0.80 "Saint Bernard"
*0.80 "St. Bernard"
* "working dogs"
* "large dogs"
* "European breeds"
$$
```

When you issue a query specifying the topic “saint-bernard”, the Enhanced Full-Text Search engine:

- Returns documents that contain one or more of the following phrases: “Saint Bernard,” “St. Bernard,” “working dogs,” “large dogs,” and “European breeds”
- Scores documents that contain the phrase “Saint Bernard” or “St. Bernard” higher than documents that contain the phrase “working dogs, “large dogs,” or “European breeds”

This example is a very basic topic definition. An outline can introduce more complex relationships by using:

- Multiple levels of subtopics
- Combinations of Verity operators (this example uses `accrue`)
Creating topics

- Verity modifiers

**Note** In Windows NT, you can use the graphical user interface of the Verity Intelligent Classifier product, which is available from Verity, to create topic outlines. If you use Intelligent Classifier, it automatically creates a topic set directory, and you can go to “Creating a knowledge base map” on page 40 to continue setting up your topics.

Creating a topic set directory

Use the mktopics utility to create and populate a topic set directory. It is located in $SYBASE/$SYBASE_FTS/verity/<verity_platform_directory>/bin.

Run, or define an alias to run, mktopics from this bin directory. You can create a topic set directory or directories in any work directory.

The mktopics syntax is:

```
mktopics -outline outline_file.otl -topicset topic_set_directory
```

Where:

- `outline_file` – is the name of the outline file you created in “Creating an outline file” on page 39
- `topic_set_directory` – is the name of the topic set directory you are creating.

For example, to execute the mktopics utility reading the saint-bernard.otl file defined above, and directing output to a work directory, use the syntax:

```
mktopics -outline /usr/u/sybase/topic_outlines/saint-bernard.otl
-topiset /usr/u/sybase/topic_sets/
saint-bernard_topic
```

Creating a knowledge base map

A knowledge base map specifies the locations of one or more topic set directories. Create an ASCII knowledge base map file that defines the fully-qualified directory paths to your topic sets.

For example, the following knowledge base map file illustrates how you can list multiple knowledge bases in the map. The first entry identifies the topic set directory created with mktopics above.
Defining the location of the knowledge base map

Set the `knowledge_base` configuration parameter to point to the location of the knowledge base map. For example:

```
sp_text_configure KRAZYKAT, 'knowledge_base', '/usr/u/sybase/topic_sets/sample_text_topics.kbm'
```

The `knowledge_base` configuration parameter is static; you must restart the Enhanced Full-Text Search engine for the definition to take effect.

Executing queries against defined topics

You can now execute queries using the defined topic instead of a complex query. For example, before you create the “saint-bernard” topic, you would have to use the following syntax:

```
...where i.index_any = "<accrue> ([80]Saint Bernard, [80]St. Bernard, working dogs, large dogs, European breeds)"
```

to find documents that:

- Contain one or more of the following phrases: “Saint Bernard,” “St. Bernard,” “working dogs,” “large dogs,” and “European breeds”
- Score documents containing the phrase “Saint Bernard” or “St. Bernard” higher than documents containing the phrase “working dogs,” “large dogs,” or “European breeds”

After you create the topic “saint-bernard”, you can use this syntax:

```
...where i.index_any = "<topic>saint-bernard"
```
or:
Creating topics

...where i.index_any = "saint bernard"

**Note** If you enter a word in a query expression, the Enhanced Full-Text Search engine tries to match it with a topic name. If you enter a phrase in a query expression, the Enhanced Full-Text Search engine replaces spaces with hyphens (-), and then tries to match it with a topic name. For example, the Enhanced Full-Text Search engine matches “saint bernard” with the topic “saint-bernard”.

See the *sample_text_topics.sql* file for examples of using topics in queries.

Troubleshooting topics

If the *knowledge_base* configuration parameter specifies a knowledge base map file that does not exist, the Enhanced Full-Text Search engine cannot start a session with Verity, and the server will not start. If the map file exists but contains invalid entries, Verity issues warning messages at start-up time. You can correct errors by editing the `<textserver>.cfg` file in the `$SYBASE` directory. You can correct path information and change the line beginning: “knowledge_base=”. 
CHAPTER 5

Writing Full-Text Search Queries

This chapter describes the pseudo columns, search operators, and modifiers that you can include in a full-text search.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components of a full-text search query</td>
<td>43</td>
</tr>
<tr>
<td>Pseudo columns in the index table</td>
<td>44</td>
</tr>
<tr>
<td>Full-text search operators</td>
<td>50</td>
</tr>
<tr>
<td>Operator modifiers</td>
<td>60</td>
</tr>
</tbody>
</table>

Components of a full-text search query

To write a full-text search query, enter the search parameters as part of an Adaptive Server `select` statement. Then the Enhanced Full-Text Search engine processes the search. The `select` statement requires:

- A `where` clause that assigns a Verity language query to the `index_any` pseudo column
- Pseudo columns to further define the parameters of the search (optional)
- A join between the IDENTITY column or primary key from the source table and the `id` column from the index table

For example, to return the 10 documents from the `copy` column of the `blurbs` table that have the most occurrences of the word “software,” enter:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id = t2.id
and t1.index_any = "<many> <word> software"
and t1.max_docs = 10
```

Adaptive Server passes the Verity query to the Enhanced Full-Text Search engine to process the search. See “How a full-text search works” on page 10.
Default behaviour

The default or simple syntax of a query to the Enhanced Full-Text Search engine is treated broadly:

1. Searches are case sensitive.
2. The STEM operator applies to search words.
3. The MANY modifier is applied.
4. The ACCRUE operator is activated at the parent level.

Pseudo columns in the index table

*Pseudo columns* are columns in the index table that define the parameters of the search and provide access to the results data. See “The index table” on page 7. These columns are valid only in the context of a query; that is, the information in the columns is valid only for the duration of the query. If the query that follows contains a different set of parameters, the pseudo columns contain a different set of values.

Each pseudo column in an index table describes a different search attribute. For example, if you indicate the score column, the query displays only the result set that falls within the parameters you define. For example, the following query displays only the results that have a score value greater than 90:

\[
\text{index_table_name.score} > 90
\]

Other pseudo columns (like highlight) are used to retrieve data generated by Verity for a particular document. Table 5-1 describes the pseudo columns that are maintained by the Enhanced Full-Text Search engine.

<table>
<thead>
<tr>
<th>Pseudo column name</th>
<th>Description</th>
<th>Datatype</th>
<th>Length (in bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster_number</td>
<td>Contains the cluster that the row is part of. Clusters are numbered starting with 1. You can use the cluster_number column only in the select clause of a query.</td>
<td>int</td>
<td>4</td>
</tr>
<tr>
<td>cluster_keywords</td>
<td>Contains the keywords that Verity uses to build the cluster. You can use cluster_keywords only in the select clause of a query.</td>
<td>varchar</td>
<td>255</td>
</tr>
<tr>
<td>highlight</td>
<td>Offsets within the document all words from the query. You can use highlight only in the select clause of a query.</td>
<td>text</td>
<td>16</td>
</tr>
</tbody>
</table>
The following sections describe the functionality of the pseudo columns.

### Using the score column to relevance-rank search results

*Relevance ranking* is the ability of the Enhanced Full-Text Search engine to assign the score parameter a value that indicates how well a document satisfies the query. The score calculation depends on the search operator used in the query. See “Using the Verity operators” on page 52. The closer the document satisfies the query, the higher the score value is for that document.

For example, if you search for documents that contain the word “rain,” a document with 12 occurrences of “rain” receives a higher score value than a document with 6 occurrences of “rain.”

---

<table>
<thead>
<tr>
<th>Pseudo column name</th>
<th>Description</th>
<th>Datatype</th>
<th>Length (in bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Uniquely identifies a document within a collection. Used to join with the IDENTITY column of the source table. You can use id in the select clause or where clause of a query.</td>
<td>numeric</td>
<td>6</td>
</tr>
<tr>
<td>index_any</td>
<td>Provides a Verity language query to the Enhanced Full-Text Search engine. You can use index_any only in a where clause. Although the pseudo-column is defined as char(255), the maximum length of the index_any clause is 16000.</td>
<td>varchar</td>
<td>255</td>
</tr>
<tr>
<td>max_docs</td>
<td>Limits results to the first n documents, based on the default sort order. In a clustered result set, limits results to the first n documents in each cluster. You can use max_docs only in a where clause.</td>
<td>int</td>
<td>4</td>
</tr>
<tr>
<td>score</td>
<td>The normalized measure of correlation between search strings and indexed columns. The score associated with a specific document has meaning only in reference to the query used to retrieve the document. You can use score in a select clause or a where clause.</td>
<td>int</td>
<td>4</td>
</tr>
<tr>
<td>sort_by</td>
<td>Specifies the sort order in which to return the result set. The Enhanced Full-Text Search engine allows up to 16 sort specifications in the sort_by column. You can use sort_by only in a where clause.</td>
<td>varchar</td>
<td>35</td>
</tr>
<tr>
<td>summary</td>
<td>Selects summarization data. You can use the summary column only in the select clause of a query.</td>
<td>varchar</td>
<td>255</td>
</tr>
<tr>
<td>total_docs</td>
<td>Contains the total number of documents that matched the search criteria.</td>
<td>int</td>
<td>4</td>
</tr>
</tbody>
</table>
If you set `score` to a high value (such as 90) in your query, you limit the result set to documents that have a `score` value greater than that number.

**Note** Verity uses decimals for `score` values; Sybase uses whole numbers. For example, if Verity reports a score value of .85, Sybase reports the same value as 85.

For example, the following query searches for documents that contain the word “raconteur” or “Paris,” or both, and that have a score of 90 or greater:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 90
and t1.index_any = "<accrue>(raconteur, Paris)"
```

<table>
<thead>
<tr>
<th>score</th>
<th>copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-----</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>(0 rows affected)</td>
<td></td>
</tr>
</tbody>
</table>

The query does not find any documents that contain the word “raconteur” or “Paris” and that have a score greater than 90. However, if the `score` value in the query is lowered to 39, you find that one document in the `blurbs` table mentions the word “raconteur” or “Paris:”

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 39
and t1.index_any = "<accrue>(raconteur, Paris)"
```

<table>
<thead>
<tr>
<th>score</th>
<th>copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-----</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>40</td>
<td>A chef's chef and a raconteur's raconteur, Reginald Blotchet-Halls calls London his second home. &quot;Th' palace</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

**Using the sort_by column to specify a sort order**

The sort order specifies the collating sequence used to order the data in the result set. The default sort order is set by the `sort_order` configuration parameter. See “Setting the default sort order” on page 72. Case insensitive sort order is supported.

Use the `sort_by` pseudo column to return a result set with a sort order other than the default. You can specify up to 16 sort specifications in the `sort_by` pseudo column.
Table 5-2 lists the values for the sort_by pseudo column.

Table 5-2: Values for the sort_by pseudo column

<table>
<thead>
<tr>
<th>Value</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>fts_score desc</td>
<td>A result set sorted by score in descending order.</td>
</tr>
<tr>
<td>fts_score asc</td>
<td>A result set sorted by score in ascending order.</td>
</tr>
<tr>
<td>fts_timestamp desc</td>
<td>A result set sorted by a timestamp in descending order.</td>
</tr>
<tr>
<td>fts_timestamp asc</td>
<td>A result set sorted by a timestamp in ascending order.</td>
</tr>
<tr>
<td>column_name desc</td>
<td>A result set sorted according to the descending order of a column.</td>
</tr>
<tr>
<td>column_name asc</td>
<td>A result set sorted according to the ascending order of a column.</td>
</tr>
<tr>
<td>fts_cluster asc</td>
<td>A clustered result set.</td>
</tr>
</tbody>
</table>

Note Before you can sort by specific columns, you must modify the style.vgw and style.ufl files. See “Setting up a column to use as a sort specification” on page 30.

For example, the following query sorts the documents by timestamp in ascending order:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 90
and t1.index_any = "<accrete>(raconteur, Paris)"
and t1.sort_by = "fts_timestamp asc"
```

Using the summary column to summarize documents

Use the summary pseudo column to have queries return only summaries of the documents that meet the search criteria, rather than returning entire documents. The summary column is not available by default; you must edit the style.prm file prior to creating the text index to enable summarization. See “Enabling query-by-example, summarization, and clustering” on page 27.

For example, the following query returns only summaries of documents that include the words “Iranian” and “book” (in this example, the style.prm file is configured to display 255 characters):

```sql
select t1.score, t1.summary
from i_blurbs t1, blurbs t2
```
Pseudo columns in the index table

where t1.id=t2.id and t1.score > 70
and t1.index_any = "(Iranian <and> book)"

<table>
<thead>
<tr>
<th>score</th>
<th>summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>They asked me to write about myself and my book, so here goes: I started a restaurant called &quot;de Gustibus&quot; with two of my fri</td>
</tr>
</tbody>
</table>

(1 row affected)

The Enhanced Full-Text Search engine supports summaries of up to 255 bytes.

For additional examples of queries using summarization, see the sample script sample_text_queries.sql in the $SYBASE/$SYBASE_FTS/sample/scripts directory.

Using pseudo columns to request clustered result sets

The clustering function analyzes a result set and groups rows into clusters so that the rows in each cluster are semantically more similar to each other, on average, than they are to other rows in other clusters. Ordering rows by subtopics can help you get a sense of the major subject areas covered in the result set.

Returning a clustered result set can be significantly slower than returning a nonclustered result set. If the response time of a query is critical, use a nonclustered result set.

Preparing to use clustering

Before you request a clustered result set, you must build the text index with the clustering function enabled. See “Enabling query-by-example, summarization, and clustering” on page 27.

The Verity clustering algorithm attempts to group similar rows together, based on the values of the following configuration parameters:

- cluster_style
- cluster_max
- cluster_effort
- cluster_order
Use `sp_text_cluster` to have a query use values that are different from the default values of these configuration parameters. See `sp_text_cluster` on page 152.

**Writing queries requesting a clustered result set**

To obtain a clustered result set, specify “fts_cluster” as the sort specification in the `sort_by` pseudo column of the query. For example:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "<many> <word> software"
and t1.max_docs = 10
and t1.sort_by = "fts_cluster asc"
```

Include any of the following pseudo columns in your query to return additional clustering information:

- **cluster_number** – contains the number of the cluster the row belongs to. Clusters are numbered starting with 1.
- **cluster_keywords** – contains the most common words found in the cluster. The `cluster_keywords` column contains a null value for each row that does not fit into any cluster.
- **max_docs** – limits the number of rows returned for each cluster. In a nonclustered query, the `max_docs` column limits the total number of rows that are returned in a result set.
- **score** – contains a value of 0 to 10000. The higher the score, the closer the row is to the cluster center. A score of 0 indicates the row does not fit into any cluster. In a nonclustered query, the `score` column can contain a value of 0 to 100. The search engine does not return results with a score of 0. A score of 0 represents “no match” but the user never sees a score of 0.

See the sample script named `sample_text_queries.sql` in the `$SYBASE/$SYBASE_FTS/sample/scripts` directory for examples of SQL statements using clustering.
Full-text search operators

The special search operators that you use to perform full-text searches are part of the Verity search engine. Table 5-3 describes the Verity search operators provided by the Enhanced Full-Text Search engine.

**Note** Sybase does not support indexing XML elements into fields.

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>accrue</td>
<td>Selects documents that contain at least one of the search elements specified in a query. The more search elements there are, the higher the score.</td>
</tr>
<tr>
<td>and</td>
<td>Selects documents that contain all the search elements specified in a query.</td>
</tr>
<tr>
<td>complement</td>
<td>Returns the complement of the score value (the score value subtracted from 100).</td>
</tr>
<tr>
<td>in</td>
<td>Selects documents that contain the search criteria in the document zone specified.</td>
</tr>
<tr>
<td>like</td>
<td>Selects documents that are similar to the sample documents or passages specified in a query.</td>
</tr>
<tr>
<td>near</td>
<td>Selects documents containing the specified search elements, where the closer the search terms are to each other in a document, the higher the document’s score.</td>
</tr>
<tr>
<td>near/n</td>
<td>Selects documents containing two or more search terms within ( n ) number of words of each other, where ( n ) is an integer up to 1000. The closer the search terms are to each other in a document, the higher the document’s score.</td>
</tr>
<tr>
<td>or</td>
<td>Selects documents that contain at least one of the search elements specified in a query.</td>
</tr>
<tr>
<td>paragraph</td>
<td>Selects documents that include all the search elements you specify within the same paragraph.</td>
</tr>
<tr>
<td>phrase</td>
<td>Selects documents that include a particular phrase. A phrase is a grouping of two or more words that occur in a specific order.</td>
</tr>
<tr>
<td>product</td>
<td>Multiplies the score values for each of the items of the search criteria.</td>
</tr>
<tr>
<td>sentence</td>
<td>Selects documents that include all the specified words in the same sentence.</td>
</tr>
<tr>
<td>stem</td>
<td>Expands the search to include the specified word and its variations.</td>
</tr>
<tr>
<td>sum</td>
<td>Adds the score values for all items in the search criteria.</td>
</tr>
<tr>
<td>thesaurus</td>
<td>Expands the search to include the specified word and its synonyms.</td>
</tr>
<tr>
<td>topic</td>
<td>Specifies that the search term you enter is a topic.</td>
</tr>
<tr>
<td>typo/n</td>
<td>Expands the search to include the specified word plus words that are similar. The optional ( n ) variable specifies the maximum number of errors between the query term and the matched term.</td>
</tr>
<tr>
<td>wildcard</td>
<td>Matches wildcard characters included in search strings. Certain characters indicate a wildcard specification automatically.</td>
</tr>
<tr>
<td>word</td>
<td>Performs a basic word search, selecting documents that include one or more instances of the specified word.</td>
</tr>
<tr>
<td>yesno</td>
<td>Converts all nonzero score values to 100.</td>
</tr>
</tbody>
</table>
Considerations when using Verity operators

When you write full-text search queries:

- Verity treats dash, underscore and ampersand as alphabetical characters rather than word separators. This behavior can be modified by changing the locale specific definition file. For example: `tokenized_as_alphabet-&` can be changed to `tokenized_as_alphabet_&`.

- You must enclose the operators in angle brackets (`<>`) in the query. If they are not enclosed in angle brackets, the Enhanced Full-Text Search engine issues error messages similar to the following:

```
Msg 20200, Level 15, State 0:
Server 'KRAZYKAT', Line 1:
Error E1-0111 (Query Builder): Syntax error in query string near character 5
Msg 20200, Level 15, State 0:
Server 'KRAZYKAT', Line 1:
Error E1-0114 (Query Builder): Error parsing query: word(tasmanian)
Msg 20101, Level 15, State 0:
Server 'KRAZYKAT', Line 1:
VdkSearchNew failed with vdk error (-40).
Msg 20101, Level 15, State 0:
Server 'KRAZYKAT', Line 1:
VdkSearchGetInfo failed with vdk error (-11).
```

- You must enclose the Verity language query in single quotes (`'`) or double quotes (`"`). The Enhanced Full-Text Search engine strips off the outermost quotes before it sends the query to Verity. For example, when you enter the query:

```
...where index_any = "'?own"
```

The Enhanced Full-Text Search engine sends the following query to Verity:

```
'?own'
```

- A query may comprise several "index_any" clauses and ed together in SQL. The right and value strings can be prefixed with "<snnn>". All such strings are concatenated in Enhanced Full-Text Search in the order determined by the "nnn" values. The "<snnn>" is removed. For instance:

```
where index_any="<s001>hello"
and index_any="<s002> world"
```
is the same as:

```sql
WHERE INDEX_ANY = "hello world"
```

This is a workaround for search strings that are greater than 255 characters.

- Search terms entered in mixed case automatically become case sensitive. Search terms entered in all uppercase or all lowercase are not automatically case sensitive. For example, a query on “Server” finds only the string “Server”. A query on “server” or “SERVER” finds the strings “Server”, “server”, and “SERVER”.

- You can use alternative syntax for the query expressions shown in Table 5-4.

<table>
<thead>
<tr>
<th>Table 5-4: Alternative Verity syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard query expression</td>
</tr>
<tr>
<td>&lt;MANY&gt;&lt;WORD&gt;string</td>
</tr>
<tr>
<td>&lt;MANY&gt;&lt;STEM&gt;string</td>
</tr>
</tbody>
</table>

When using the alternative syntax, remember that the Enhanced Full-Text Search engine strips off the outermost quotes before it sends the query to Verity. For example, when you enter the query:

```sql
...where index_any = "'play'"
```

the Enhanced Full-Text Search engine sends the following query to Verity:

```sql
'play'
```

This is the same as:

```sql
<MANY><STEM>play
```

### Using the Verity operators

The following sections describe how to use the Verity operators shown in Table 5-3 on page 50.

#### accrue

The `accrue` operator selects documents that contain at least one of the search items specified in the query. There must be two or more search elements. Each result is relevance-ranked. For example, the following query searches for the word “restaurant” or “deli” or both in the `copy` column of the `blurbs` table:
and, or

The and and or operators select documents that contain the specified search elements. Each result is relevance-ranked. The and operator selects documents that contain all the elements specified in the query. For example, the following query selects documents that contain both “Iranian” and “business”:

```sql
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "(Iranian <and> business)"
```

The or operator selects the documents that contain any of the search elements. For example, if the preceding query is rewritten to use the or operator, the query selects documents that contain the word “Iranian” or “business”:

```sql
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "(Iranian <or> business)"
```

complement

The complement operator returns the complement of the score value for a document; that is, it subtracts the value of score from 100 and returns the result as the score value for the document.

in

The in operator selects documents that contain the specified search element in one or more document zones. Document zones are created for a text index in the following two scenarios:

- When you create an index on two or more columns using `sp_create_text_index`, a document zone is created for each column in the text index. See “Specifying multiple columns when creating a text index” on page 22. A document zone is not created when you create a text index on a single column. For example, if you specify the `au_id` and `copy` columns of the `blurbs` table when you create the text index, you can issue the following query:
**Full-text search operators**

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 35
  and t1.index_any = "gorilla <in> copy"
```

This returns rows that contain the word “gorilla” in the `copy` column. However, if you specify only the `copy` column of the `blurbs` table when you create the text index, this query does not return any rows.

- When you create an index that uses a filter, a document zone is created for each tag in the document. See “Using filters on text that contains tags” on page 32. You can limit your search to a particular tag by specifying the tag name after the `in` operator. For example, to search for the word “automotive” in a “title” tag in an HTML document, specify:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 35
  and t1.index_any = "automotive <in> title"
```

Text indexes utilizing filters can contain only one column.

### like

The `like` operator selects documents that are similar to the documents or passages you provide. The search engine analyzes the text to find the most important terms to use. If you specify multiple samples, the search engine selects important terms that are common across the samples. Each result is relevance-ranked.

The `like` operator accepts a single operand, called the query-by-example (QBE) specification. The QBE specification can be either literal text or document IDs. The document IDs are from the IDENTITY column in the source table. For example, to select documents that are similar to the document in the `copy` column in the row with an IDENTITY of “2”, enter:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 35
  and t1.index_any = '<like> ( "{2}" )'
```

Before using literal text in the QBE specification, you must uncomment the following line in the `style.prm` file:

```bash
$define DOC-FEATURES "TF"
```

See “Enabling query-by-example, summarization, and clustering” on page 27.
See the sample script named `sample_text_queries.sql` in the `$SYBASE/$SYBASE_FTS/sample/scripts` directory for examples of SQL statements using QBE.

**near, near/n**

The `near` operator selects documents that contain the items specified in the query and are near each other ("near" being a relative term). The documents in which the search words appear closest to each other receive the highest relevance-ranking.

The `near/n` operator specifies how far apart the items can be (n has a maximum value of 1000). The following example selects documents in which the words "raconteur" and "home" appear within 10 words of each other:

```sql
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "<near/10>(raconteur, home)"
```

**or**

See “and, or” on page 53.

**phrase**

The `phrase` operator selects documents that contain a particular phrase (a group of two or more items that occur in a specific order). Each result is relevance-ranked. The following example selects the documents that contain the phrase "gorilla's head":

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 50
and t1.index_any = "<phrase>(gorilla's head)"
```

**paragraph**

The `paragraph` operator selects documents in which the specified search elements appear in the same paragraph. The closer the words are to each other in a paragraph, the higher the score the document receives in relevance-ranking. The following example searches for documents in which the words "text" and "search" occur within the same paragraph:
Full-text search operators

    select t1.score, t2.copy
    from i_blurbs t1, blurbs t2
    where t1.id=t2.id and t1.score > 50
    and t1.index_any = "<many><paragraph>(text, search)"

**product**

The product operator multiplies the score value for the documents for each of the search elements. To arrive at a document’s score, the Enhanced Full-Text Search engine calculates a score for each search element and multiplies the scores. For example:

    select t1.score, t2.copy
    from i_blurbs t1, blurbs t2
    where t1.id=t2.id and t1.score > 50
    and t1.index_any = "<product>(cat, created)"

The score value for each search element is 78; however, because the score values for the items are multiplied, the document has a score value of 61 (.78 x .78 = .61(100) = 61).

**sentence**

The sentence operator selects documents in which the specified search elements appear in the same sentence. The closer the words are to each other in a sentence, the higher the score the document receives in relevance-ranking. The following example searches for documents in which the words “tax” and “service” occur within the same sentence:

    select t1.score, t2.copy
    from i_blurbs t1, blurbs t2
    where t1.id=t2.id and t1.score > 50
    and t1.index_any = "<many><sentence>(tax, service)"

**stem**

The stem operator searches for documents containing the specified word and its variations. For example, if you specify the word “cook,” the Enhanced Full-Text Search engine produces documents that contain “cooked,” “cooking,” “cooks,” and so on. To relevance-rank the result set, include the many modifier in the query. See “Operator modifiers” on page 60.
The following query uses the stem operator to find documents that contain variations of the word “create,” that is, words that contain the word “create” as a stem. Notice that even though the first document contains a word in which “create” is not a perfect stem (“creative”), the document is still selected:

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 10
and t1.index_any = "<many><stem>create"
```

78 Anne Ringer ran away from the circus as a child. A university creative writing professor and her family . . .

78 If Chastity Locksley didn't exist, this troubled world would have created her! Not only did she master the mystic

**sum**

The `sum` operator totals the score values for each search element, up to a maximum of 100. To arrive at a document’s score, the Enhanced Full-Text Search engine calculates a score for each search element and totals those scores.

**thesaurus**

The `thesaurus` operator searches for documents containing a synonym for a search element. For example, you might perform a search using the word “dog,” looking for documents that use any of its synonyms (“canine,” “pooch,” “pup,” “watchdog,” and so on). Each result is relevance-ranked.

The Enhanced Full-Text Search engine supplies a default thesaurus. You can also create a custom thesaurus. See “Creating a custom thesaurus” on page 34.

The following example uses the `thesaurus` operator to find a result set that contains synonyms for the word “crave.” The first document is selected because it contains the word “want;” the second, because it contains the word “hunger:”

```sql
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "<thesaurus>(crave)"
```

score copy
They asked me to write about myself and my book, so here goes: I started a restaurant called "de Gustibus" with two of restaurant over another, when what they really want is a... A chef's chef and a raconteur's raconteur, Reginald Blotchet-Halls calls London his second home. "Th' palace... his equal skill in satisfying our perpetual hunger for... topic

The topic operator selects documents that meet the search criteria defined by the specified topic. See “Creating topics” on page 38. For example, use the following syntax to find documents that meet the criteria defined by the topic “engineering:"

```
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = "<topic>{engineering}" 
```

typo/n

The typo/n operator expands the search to include the specified word plus words that are similar. The optional n variable specifies the maximum number of errors between the query and the matched terms. For example, typo mouse returns documents that include words such as “house,” “louse,” and “moose.”

wildcard

The wildcard operator allows you to substitute wildcard characters for part of the item for which you are searching. Table 5-5 describes the wildcard characters and their attributes.

<table>
<thead>
<tr>
<th>Character</th>
<th>Function</th>
<th>Syntax</th>
<th>Locates</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>Specifies one alphanumeric character. You do not need to include the wildcard operator when you include the question mark in your query. The question mark is ignored in a set ([]) or in an alternative pattern ({}).</td>
<td>?an'</td>
<td>“ran,” “pan,” “can,” and “ban”</td>
</tr>
</tbody>
</table>
To relevance-rank the result set, include the many modifier in the query. See “Operator modifiers” on page 60.

For example, the following query searches for documents that include variations of the word “slingshot”:

```
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = 'slingshot*'
score copy
```

100 Albert Ringer was born in a trunk to circus parents, but another kind of circus trunk played a more important role

. . .
gorilla. "Slingshotting" himself from the ring ropes,

. . .

**word**

The word operator searches for documents containing the specified word. To relevance-rank the result set, include the many operator in the query. The following example searches the blurbs table for documents containing the word “palates”:

```
select t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id
and t1.index_any = 'palates'
score copy
```

100 Albert Ringer was born in a trunk to circus parents, but another kind of circus trunk played a more important role

. . .
gorilla. "Slingshotting" himself from the ring ropes,
**Operator modifiers**

```sql
select t1.score, t2.copy
from i_blurbs t1, blurbs t2
where t1.id=t2.id and t1.score > 50
and t1.index_any = "<many><word>(palates)"
```

**yesno**

The yesno operator converts all nonzero score values to 100. For example, if the score values for five documents are 86, 45, 89, 89, and 100, each of those documents is returned with a score value of 100. Score values of 0 are not changed. The yesno operator is helpful for ensuring that all documents containing the search criteria are returned in the result set, regardless of the sort order.

**Operator modifiers**

The Verity query language includes modifiers that you can use with the operators to refine a search. The modifiers are described in Table 5-6.

**Table 5-6: Verity operator modifiers**

<table>
<thead>
<tr>
<th>Modifier name</th>
<th>Description</th>
<th>Works with these operators</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>case</td>
<td>Performs case-sensitive searches. If you enter search terms in mixed case, the search is automatically case sensitive.</td>
<td>wildcard, word</td>
<td><code>&lt;case&gt;&lt;word&gt;(Net)</code></td>
</tr>
<tr>
<td>many</td>
<td>Counts the number of times that a word, stemmed word, or phrase occurs in a document. Relevance-ranks the document according to its density.</td>
<td>paragraph, phrase, sentence, stem, word, wildcard</td>
<td><code>&lt;many&gt;&lt;stem&gt;(write)</code></td>
</tr>
<tr>
<td>not</td>
<td>Excludes documents that contain the items for which the query is searching.</td>
<td>and, or</td>
<td><code>cat&lt;and&gt;&lt;not&gt;elephant</code></td>
</tr>
</tbody>
</table>
CHAPTER 5  Writing Full-Text Search Queries

<table>
<thead>
<tr>
<th>Modifier name</th>
<th>Description</th>
<th>Works with these operators</th>
<th>Example</th>
</tr>
</thead>
</table>
| order         | Specifies that the items in the documents occur in the same order in which they appear in the query. Always place the order modifier just before the operator. | near/n paragraph sentence | Simple syntax: tidbits<order><paragraph>king  
Explicit syntax: <order><paragraph>(tidbits,king) |
This chapter describes system administration issues for the Enhanced Full-Text Search engine.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting the Enhanced Full-Text Search engine on UNIX</td>
<td>63</td>
</tr>
<tr>
<td>Starting the Enhanced Full-Text Search engine on Windows NT</td>
<td>65</td>
</tr>
<tr>
<td>Shutting down the Enhanced Full-Text Search engine</td>
<td>66</td>
</tr>
<tr>
<td>Modifying the configuration parameters</td>
<td>67</td>
</tr>
<tr>
<td>Backup and recovery for the Enhanced Full-Text Search engine</td>
<td>75</td>
</tr>
</tbody>
</table>

**Starting the Enhanced Full-Text Search engine on UNIX**

Use the `startserver` utility to start the Enhanced Full-Text Search engine on UNIX. The `startserver` utility is included in the `install` directory of Adaptive Server. For example, to start a Enhanced Full-Text Search engine named KRAZYKAT, enter:

```
startserver -f
$SYBASE/$SYBASE_FTS/install/RUN_KRAZYKAT
```

where the `-f` flag specifies the relative path to the runserver file. After you issue the command, the Enhanced Full-Text Search engine issues a series of messages describing the settings of the configuration parameters.

**Creating the runserver file**

The runserver file contains start-up commands for the Enhanced Full-Text Search engine. The `runserver` file can include these flags:

- `-server_name` – specifies the name of the Enhanced Full-Text Search engine and is used to locate the configuration file and the network connection information in the interfaces file.
Starting the Enhanced Full-Text Search engine on UNIX

- `-t` – causes the Enhanced Full-Text Search engine to write start-up messages to standard error.
- `-errorlog_path` – specifies the path to the error log file.
- `-interfaces_file_path` – specifies the path to the interfaces file.

A sample runserver file is copied to the $SYBASE/$SYBASE_FTS/install directory during installation. Make a copy of this file, renaming it `RUN_server_name`, where `server_name` is the name of the Enhanced Full-Text Search engine. You must include the correct path environment variable for your platform in the runserver file. Table 6-1 shows the path environment variable to use for each platform.

### Table 6-1: Path environment variable for the runserver file

<table>
<thead>
<tr>
<th>Platform</th>
<th>Environment variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS/6000 AIX</td>
<td>LIBPATH</td>
</tr>
<tr>
<td>Sun Solaris</td>
<td>LD_LIBRARY_PATH</td>
</tr>
<tr>
<td>HP 9000(800)</td>
<td>SHLIB_PATH</td>
</tr>
<tr>
<td>Compaq Tru64</td>
<td>LD_LIBRARY_PATH</td>
</tr>
<tr>
<td>Linux</td>
<td>LD_LIBRARY_PATH</td>
</tr>
<tr>
<td>NT</td>
<td>PATH</td>
</tr>
</tbody>
</table>

For example, the runserver file on Sun Solaris for an Enhanced Full-Text Search engine named KRAZYKAT would be `RUN_KRAZYKAT` and would be similar to:

```
#!/bin/sh
#

LD_LIBRARY_PATH=$SYBASE/$SYBASE_FTS/lib:$LD_LIBRARY_PATH
export LD_LIBRARY_PATH

$SYBASE/bin/txtsvr -SKRAZYKAT
```

The start-up command in the runserver file must consist of a single line and cannot include a return. If you must carry the contents of the file over to a second or third line, include a backslash (\) for a line break.
Starting the Enhanced Full-Text Search engine on Windows NT

You can start the Enhanced Full-Text Search engine from Sybase Central™, as a service, or from the command line:

- From Sybase Central – see your Sybase Central documentation for information about starting servers.
- As a service – see “Starting the Enhanced Full-Text Search engine as a service” on page 65
- From the command line – use the following syntax:

  \%SYBASE\%SYBASE_FTS\%bin\txtsvr.exe -S server_name
  \[-t\] \[-i\%SYBASE\%path_to_sql.ini_file\] \[-l\%SYBASE\%path_to_errorlog\]

Where:

- `-S` is the name of the Enhanced Full-Text Search engine you are starting.
- `-t` directs start-up messages to standard error.
- `-i` is the path to the `sql.ini` file. This path must include the `sql.ini` file name.
- `-l` is the path to the error log. This path must include the name of the error log file.

For example, to start a Enhanced Full-Text Search engine named KRAZYKAT on NT using the default `sql.ini` and error log files, and using `-t` to trace start-up messages, enter:

\%SYBASE\%SYBASE_FTS\%bin\txtsvr.exe -SKRAZYKAT -t

The Enhanced Full-Text Search engine is running when you see the start-up complete message.

Starting the Enhanced Full-Text Search engine as a service

Use the `instsvr` utility in Sybase Central to add the Enhanced Full-Text Search engine to the list of items you can start and stop with the Services utility. `instsvr` is located in the `%SYBASE\%SYBASE_FTS\%bin` directory.

The `instsvr` utility uses the following syntax:

```
instsvr.exe service_name %SYBASE\%SYBASE_FTS\%bin\txtsvr.exe
```

65
Shutting down the Enhanced Full-Text Search engine

"startup_parameters"

where:

- **service_name** – is the name of the Enhanced Full-Text Search engine you are adding as a service. With Sybase Central, Sybase recommends you use a server name with the extension "_TS" (for example, KRAZYKAT_TS).

- **startup_parameters** – are any parameters you want used at start-up.

For example, to install an Enhanced Full-Text Search engine named KRAZYKAT_TS as a service, enter:

```
instsvr.exe KRAZYKAT_TS %SYBASE%\sds\text\bin\txtsvr.exe
"-SKRAZYKAT_TS -t"
```

**Note** To include more than one parameter (for example, `-i`), enclose all the parameters in one set of double quotes.

To configure Sybase Central to start and stop your Enhanced Full-Text Search engine, you must provide a service name that begins with "SYBTXT_server_name" where **server_name** is the name of the Enhanced Full-Text Search engine listed in the interfaces file. For example, if the name in the interfaces file is KRAZYKAT_TS, run the following instsvr command to create a service that can be managed by Sybase Central:

```
instsvr SYBTXT_KRAZYKAT_TS %SYBASE%\%SYBASE_FTS%\bin\txtsvr.exe
"-SKRAZYKAT_TS -t"
```

**Shutting down the Enhanced Full-Text Search engine**

Use the following command to shut down the Enhanced Full-Text Search engine from Adaptive Server:

```
server_name...sp_shutdown
```

where **server_name** is the name of the Enhanced Full-Text Search engine you are shutting down.

Only users with sa_role can shut down the Enhanced Full-Text Search engine.

For example, to shutdown a Enhanced Full-Text Search engine named KRAZYKAT, enter:

```
KRAZYKAT...sp_shutdown
```
Modifying the configuration parameters

Each Enhanced Full-Text Search engine has configuration parameters with default values, as shown in Table 6-2.

Table 6-2: Configuration parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch_size</td>
<td>Determines the size of the batches sent to the Enhanced Full-Text Search engine.</td>
<td>500</td>
</tr>
<tr>
<td>batch_blocksize</td>
<td>When enabled, the text server reads data in smaller chunks. This parameter instructs the text server to retrieve n number of rows at a time. Range is 0 (disabled) to 65535.</td>
<td>0</td>
</tr>
<tr>
<td>max_indexes</td>
<td>The maximum number of text indexes that are created in the Enhanced Full-Text Search engine.</td>
<td>126</td>
</tr>
<tr>
<td>max_stacksize</td>
<td>Size (in kilobytes) of the stack allocated for client threads.</td>
<td>34,816</td>
</tr>
<tr>
<td>max_threads</td>
<td>Maximum number of threads available for the Enhanced Full-Text Search engine.</td>
<td>50</td>
</tr>
<tr>
<td>max_packetsize</td>
<td>Packet size sent between the Enhanced Full-Text Search engine and the Adaptive Server.</td>
<td>2048</td>
</tr>
<tr>
<td>max_sessions</td>
<td>Maximum number of sessions for the Enhanced Full-Text Search engine.</td>
<td>100</td>
</tr>
<tr>
<td>min_sessions</td>
<td>Minimum number of sessions for the Enhanced Full-Text Search engine.</td>
<td>10</td>
</tr>
<tr>
<td>language</td>
<td>Language used by the Enhanced Full-Text Search engine.</td>
<td>us_english</td>
</tr>
<tr>
<td>charset</td>
<td>Character set used by the Enhanced Full-Text Search engine.</td>
<td>iso_1</td>
</tr>
<tr>
<td>vdkCharset</td>
<td>Character set used by Verity search engine.</td>
<td>850</td>
</tr>
<tr>
<td>vdkLanguage</td>
<td>Language used by Verity search engine.</td>
<td>english</td>
</tr>
<tr>
<td>vdkHome</td>
<td>Verity directory.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UNIX: $SYBASE/$SYBASE_FTS/verity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Windows NT: %SYBASE%%SYBASE_FTS\verity</td>
<td></td>
</tr>
<tr>
<td>collDir</td>
<td>Storage location of the Enhanced Full-Text Search engine’s collection.</td>
<td>UNIX: $SYBASE/$SYBASE_FTS/collections</td>
</tr>
</tbody>
</table>
# Modifying the configuration parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>defaultDb</td>
<td>Name of the Enhanced Full-Text Search engine database that stores text index metadata.</td>
<td>text_db</td>
</tr>
</tbody>
</table>
| interfaces    | Full path to the directory in which the interfaces file used by the Enhanced Full-Text Search engine is located. | UNIX: $SYBASE/interfaces  
Windows NT: %SYBASE%\ini\sql.ini |
| sort_order    | Default sort order.                                                         | 0                                       |
| errorLog      | Full path name to the error log file. The directory in which you start Enhanced Full-Text Search engine |                                          |
| traceflags    | String containing numeric identifiers used to generate diagnostic information. | 0                                       |
| srv_traceflags| String containing numeric flag identifiers used to generate Open Server diagnostic information. | 0                                       |
| max_session_fd| The maximum number of file descriptors used by an Enhanced Full-Text Search session. See “File Descriptors and Enhanced Full-Text Search” on page 86. | 0                                       |
| cluster_style | Clustering style to use.                                                    | Fixed                                   |
| cluster_max   | Maximum number of clusters to generate when cluster_style is set to Fixed.  | 0                                       |
| cluster_effort| Amount of effort the Enhanced Full-Text Search engine should expend on finding a good cluster. | Default                                 |
| cluster_order | The order to return clusters and rows within a cluster.                     | 0                                       |
| auto_online   | Specifies whether to bring indexes online automatically when the Enhanced Full-Text Search engine is started. 0 indicates online is not automatic; 1 indicates automatic. | 0                                       |
| backDir       | The default location for the placement of text index backup files.           | UNIX: $SYBASE/$SYBASE_FTS/backup  
Windows NT: %SYBASE\%\%SYBASE_FTS%\backup |
| knowledge_base| The location of a knowledge base map for implementing the Verity topics feature. | null                                    |
| nocase        | Sets the case-sensitivity of the Enhanced Full-Text Search engine. If you are using a case-sensitive sort order in Adaptive Server, set to 0. If you are using a case-insensitive sort order, set to 1. | 0                                       |
A sample configuration file that includes all of these parameters is copied to your installation directory during installation. The sample configuration file, which is named `textsvr.cfg` is listed in Appendix B, “Sample Files.”

**Modifying configuration values**

Use `sp_text_configure` to change the value of a configuration parameter. The syntax is:

```
sp_text_configure server_name, config_name, config_value
```

Where:

- `server_name` – is the name of the Enhanced Full-Text Search engine.
- `config_name` – is the name of the configuration parameter.
- `config_value` – is the value you assign to the configuration parameter.

See `sp_text_configure` on page 154.

**Available configuration parameters**

Table 6-3 provides a list of available configuration parameters with valid limits:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
<th>Static/dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch_size</td>
<td>0 – MAX_INT</td>
<td>Dynamic</td>
</tr>
<tr>
<td>batch_blocksize</td>
<td>0 – 65535</td>
<td>Dynamic</td>
</tr>
<tr>
<td>max_indexes</td>
<td>0 – MAX_INT</td>
<td>Static</td>
</tr>
<tr>
<td>max_stacksize</td>
<td>0 – MAX_INT</td>
<td>Static</td>
</tr>
<tr>
<td>max_threads</td>
<td>0 – MAX_INT</td>
<td>Static</td>
</tr>
<tr>
<td>max_packetsize</td>
<td>0 – MAX_INT</td>
<td>Static</td>
</tr>
<tr>
<td>max_sessions</td>
<td>0 – MAX_INT</td>
<td>Static</td>
</tr>
<tr>
<td>min_sessions</td>
<td>0 – max_sessions</td>
<td>Static</td>
</tr>
<tr>
<td>language</td>
<td>french, spanish, german, us_english</td>
<td>Static</td>
</tr>
<tr>
<td>charset</td>
<td>ascii_8, cp037, cp1047, cp437, cp500, cp850, deckanji, eucjis, iso_1, mac, roman8, sjis, utf8</td>
<td>Static</td>
</tr>
<tr>
<td>vdkCharset</td>
<td>50, 437, 1252, mac1 (Just the ones listed in the manual)</td>
<td>Static</td>
</tr>
</tbody>
</table>
Modifying the configuration parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
<th>Static/dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>vdkLanguage</td>
<td>frenchx, spanischx, germanx, english, englishx, bokmalx, dutchx, finnischx, nynorskkx, swedishx, portugx, italianx, danishx</td>
<td>Static</td>
</tr>
<tr>
<td>vdkHome</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>collDir</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>default_Db</td>
<td>A string &lt; 32 chars</td>
<td>Static</td>
</tr>
<tr>
<td>interfaces</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>sort_order</td>
<td>0, 1, 2, 3</td>
<td>Dynamic</td>
</tr>
<tr>
<td>errorLog</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>traceflags</td>
<td>A string with comma delimited numbers ranging anywhere from 1 to 15</td>
<td>Static</td>
</tr>
<tr>
<td>srv_traceflags</td>
<td>A string with comma delimited numbers ranging anywhere from 1 to 8</td>
<td>Static</td>
</tr>
<tr>
<td>cluster_style</td>
<td>Coarse, Medium, Fine, Fixed</td>
<td>Dynamic</td>
</tr>
<tr>
<td>cluster_max</td>
<td>0 – MAX_INT</td>
<td>Dynamic</td>
</tr>
<tr>
<td>cluster_effort</td>
<td>Low, Medium, High, Default</td>
<td>Dynamic</td>
</tr>
<tr>
<td>cluster_order</td>
<td>0 or 1</td>
<td>Dynamic</td>
</tr>
<tr>
<td>auto_online</td>
<td>0 or 1</td>
<td>Static</td>
</tr>
<tr>
<td>backCmd</td>
<td>A string &lt; 255 chars</td>
<td>Dynamic</td>
</tr>
<tr>
<td>restoreCmd</td>
<td>A string &lt; 255 chars</td>
<td>Dynamic</td>
</tr>
<tr>
<td>backDir</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>knowledge_base</td>
<td>A string &lt; 255 chars</td>
<td>Static</td>
</tr>
<tr>
<td>nocase</td>
<td>0 or 1</td>
<td>Dynamic</td>
</tr>
<tr>
<td>max_session_fd</td>
<td>0, 5 – MAX_INT</td>
<td>Static</td>
</tr>
</tbody>
</table>

Setting the default language

The default language for Verity is set with the vdkLanguage configuration parameter. By default, vdkLanguage is set to “english.” You can configure Verity to use a different default language. Table 6-4 lists the locales supported by Sybase.
Table 6-4: vdkLanguage configuration parameters

<table>
<thead>
<tr>
<th>Language</th>
<th>Default locale name</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>english</td>
</tr>
<tr>
<td>German</td>
<td>german</td>
</tr>
<tr>
<td>French</td>
<td>french</td>
</tr>
</tbody>
</table>

Additional language adapters are available in the `SYBASE/SYBASE_FTS/verity/common` directory; however, the Enhanced Full-Text Search engine displays messages only in the languages shown in Table 6-4.

The language parameter is the language the Enhanced Full-Text Search engine uses to display its error messages and Open Server and Open Client error messages. Set the language parameter to the Adaptive Server language.

With the Enhanced Full-Text Search engine, run:

```
sp_text_configure KRAZYKAT, 'vdkLanguage', 'spanish'
```

For more information about the Verity languages, see the Verity Web site at http://www.verity.com.

Setting the default character set

Set the default character set for Verity using the `vdkCharset` parameter in the configuration file. The files used for the Verity character sets are in `SYBASE/SYBASE_FTS/verity/common`. Table 6-5 describes the character sets you can use with Verity.

Table 6-5: Verity character sets

<table>
<thead>
<tr>
<th>Character set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>850</td>
<td>Default</td>
</tr>
<tr>
<td>437</td>
<td>IBM PC character set</td>
</tr>
<tr>
<td>1252</td>
<td>Windows code page for Western European languages</td>
</tr>
<tr>
<td>mac1</td>
<td>Macintosh roman</td>
</tr>
</tbody>
</table>

The default character set for the Enhanced Full-Text Search engine is set with the `charset` parameter. Set the `charset` parameter to the Adaptive Server character set.

For example, with the Enhanced Full-Text Search engine, run:

```
sp_text_configure KRAZYKAT, 'vdkCharset', '437'
```
Modifying the configuration parameters

Indexing on the euro symbol

To index the euro symbol, install the utf8 charset in Adaptive Server. Set vdkLanguage to `<language>x` and leave vdkCharset blank. For example:

```
ASE 12.5.x charset = utf8
EFTS 12.5.x vdkLanguage = englishx
EFTS 12.5.x vdkCharset =
```

Setting the default sort order

By default, the Enhanced Full-Text Search engine sorts the result set by the score pseudo column in descending order (the higher scores appear first). To change the default sort order, set the sort_order configuration parameter to one of the values in Table 6-6.

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Returns result sets sorted by the score pseudo column in descending order.</td>
</tr>
<tr>
<td></td>
<td>The default value.</td>
</tr>
<tr>
<td>1</td>
<td>Returns result sets sorted by the score pseudo column in ascending order.</td>
</tr>
<tr>
<td>2</td>
<td>Returns result sets sorted by a timestamp in descending order.</td>
</tr>
<tr>
<td>3</td>
<td>Returns result sets sorted by a timestamp in ascending order.</td>
</tr>
</tbody>
</table>

For example, enter:

```
sp_text_configure KRAZYKAT, 'sort_order', '2'
```

When you sort a result set by descending timestamp (value 2 in Table 6-6), the Enhanced Full-Text Search engine returns the newest documents first. The newest documents are those that were inserted or updated most recently. When results are sorted by ascending timestamp (value 3 in Table 6-6), the Enhanced Full-Text Search engine returns the oldest documents first.

Setting the default sort order is especially important if your query uses the max_docs pseudo column. The max_docs pseudo column limits the number of rows of the result set to the first n rows, ordered by the sort order. If you set max_docs to a number smaller than the size of the result set, the sort order you select could exclude the rows that contain the information for which you are searching.
For example, if you sort by ascending timestamp, the latest document added to the table appears last in the result set. If the entire result set consists of 11 documents, and you set `max docs` to 10, the latest document does not appear in the result set. However, if you sort by descending timestamp, the latest document appears first in the result set.

### Setting trace flags

The `traceflags` parameter enables the logging of certain events when they occur within the Enhanced Full-Text Search engine. Each trace flag is uniquely identified by a number. Trace flags are described in Table 6-7.

<table>
<thead>
<tr>
<th>Trace flag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Traces connects, disconnects, and attention events from Adaptive Server.</td>
</tr>
<tr>
<td>2</td>
<td>Traces language events. Traces the SQL statement that Adaptive Server sent to the Enhanced Full-Text Search engine.</td>
</tr>
<tr>
<td>3</td>
<td>Traces RPC events.</td>
</tr>
<tr>
<td>4</td>
<td>Traces cursor events. Traces the SQL statement sent to the Enhanced Full-Text Search engine by Adaptive Server.</td>
</tr>
<tr>
<td>5</td>
<td>Writes the errors that display to the log.</td>
</tr>
<tr>
<td>6</td>
<td>Traces information about text indexes. Writes the search string being passed to Verity to the log, and writes the number of records that the search returns to the log.</td>
</tr>
<tr>
<td>7</td>
<td>Traces done packets.</td>
</tr>
<tr>
<td>8</td>
<td>Traces calls to the interface between the Enhanced Full-Text Search engine and the Verity API.</td>
</tr>
<tr>
<td>9</td>
<td>Traces SQL parsing.</td>
</tr>
<tr>
<td>10</td>
<td>Traces Verity processing.</td>
</tr>
<tr>
<td>11</td>
<td>Disables Verity collection optimization.</td>
</tr>
<tr>
<td>12</td>
<td>Disables <code>sp_statistics</code> from returning information.</td>
</tr>
<tr>
<td>13</td>
<td>Traces backup operations.</td>
</tr>
<tr>
<td>14</td>
<td>Logs Verity status and timing information.</td>
</tr>
<tr>
<td>15</td>
<td>Generates ngram index information for collections. ngrams increase the speed of wildcard searches. This trace flag is required for wildcard searches against data in Unicode format.</td>
</tr>
</tbody>
</table>
Modifying the configuration parameters

You can enable and disable trace flags interactively, using the remote procedure calls (RPCs) `sp_traceon` and `sp_traceoff` in the Enhanced Full-Text Search engine.

To execute `sp_traceon`, use the following, where `textserver` is the name of the Enhanced Full-Text Search engine:

```
textserver...sp_traceon 1,2,3,4
```

The traceflags stay active until the session is terminated or until the `sp_traceoff` RPC is executed using the specific traceflag. To set a traceflag permanently, either set it in the configuration file or use the `sp_text_configure` command.

### Setting Open Server trace flags

Use the `srv_traceflags` parameter to turn on trace flags to log Open Server diagnostic information. Open Server trace flags are described in Table 6-8.

**Table 6-8: Open Server trace flags**

<table>
<thead>
<tr>
<th>Trace flag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Traces TDS headers</td>
</tr>
<tr>
<td>2</td>
<td>Traces TDS data</td>
</tr>
<tr>
<td>3</td>
<td>Traces attention events</td>
</tr>
<tr>
<td>4</td>
<td>Traces message queues</td>
</tr>
<tr>
<td>5</td>
<td>Traces TDS tokens</td>
</tr>
<tr>
<td>6</td>
<td>Traces Open Server events</td>
</tr>
<tr>
<td>7</td>
<td>Traces deferred event queues</td>
</tr>
<tr>
<td>8</td>
<td>Traces network requests</td>
</tr>
</tbody>
</table>

For example:

```
sp_text_configure KRAZYKAT, 'srv_traceflags', '3'
```
Setting case sensitivity

By default, the Enhanced Full-Text Search engine is case sensitive. This means you must enter identifiers in the same case or they are not recognized. For example, if you have a table named blurs (lowercase), you cannot issue an `sp_create_text_index` command that specifies the table name BLURBS. You must issue a command that uses the same case for the table name argument:

```sql
sp_create_text_index "KRAZYKAT", "i_blurbs", "blurbs", ",", "copy"
```

With Enhanced Full-Text Search engine, use the `nocase` parameter to set the case sensitivity of the Enhanced Full-Text Search engine. 0 indicates case sensitive; 1 indicates case insensitive. Set the `nocase` parameter to the sort order case sensitivity in Adaptive Server.

This example changes the KRAZYKAT server to case insensitive.

```sql
sp_text_configure KRAZYKAT, 'nocase', '1'
```

**Note** The `nocase` parameter does not affect the case sensitivity of the Verity query. See “Considerations when using Verity operators” on page 51.

Backup and recovery for the Enhanced Full-Text Search engine

Backup and recovery for the Enhanced Full-Text Search Specialty Data Store is automated with `sp_text_dump_database` and `sp_text_load_index`. These system procedures provide a seamless interface for maintaining data and text index integrity.

The Adaptive Server user database and the Verity collections are physically separate. Backing up your user database does not back up the Verity collections, and restoring your database from a backup does not restore your Verity collections. The backup and recovery procedures described in Chapter 27, “Backing Up and Restoring User Databases,” of the *System Administration Guide* apply only to the user database and the `text_db` database in Adaptive Server.
Follow the recommended schedule for backing up your databases, as described in Chapter 26, “Developing a Backup and Recovery Plan,” of the System Administration Guide. Sybase recommends that when you back up a user database with text indexes, you also back up:

- The text_db database
- The text indexes

**Note** You do not have to back up the user database and text indexes at the same time to recover the text indexes. However, you must restore the user database before you restore the text index. This restores the text_events table, which sp_text_load_index uses to bring the text indexes in sync with the user database.

A regular backup schedule ensures the integrity of the text indexes, the Adaptive Server data, and the text_events table, all of which are integral to recovering your text indexes without having to drop and re-create them.

### Customizable backup and restore

tBackCmd and restoreCmd allow customizable backup and restore commands to be used instead of tar or zip commands when backing up collection files. If backCmd and restoreCmd are blank, the default commands are used; otherwise, the specified command is executed. String substitution is performed before execution to allow specification of input and output directories and collection identification. The string substitution is defined as follows:

- `{backDir}` is replaced by the `backup` directory specified as the “backDir” configuration parameter.
- `{collDir}` is replaced by the full path name for the collection.
- `{colID}` is replaced by the collection ID which is the full name of the backup file.

### Backing up Verity collections

The `sp_text_dump_database` system procedure backs up collections and (optionally) the user and text_db databases. `sp_text_dump_database` also maintains the text_events table by deleting entries that are no longer needed for recovery.
During a backup, the Enhanced Full-Text Search engine processes queries, but defers any update requests until the backup is complete. This eliminates the need to shut down and restart the Enhanced Full-Text Search engine.

Run `sp_text_dump_database` from the database containing the text indexes you are backing up. Make sure all the required servers are running when issuing `sp_text_dump_database`. `sp_text_dump_database` unconditionally backs up all indexes of all enhanced text servers. The backup of the text indexes is placed in the directory specified in the `backDir` configuration parameter. The output of `dump database` is written to the Enhanced Full-Text Search error log. Sybase recommends dumping the current database and the `text_db` database at the time the text indexes are backed up. However, this is optional.

For example, to back up the text indexes, the `sample_colors_db` database to the `/work2/sybase/colorsbackup` directory, and the `text_db` database to the `/work2/sybase/textdbbackup` directory, enter:

```
sp_text_dump_database @backupdbs = INDEXES_AND_DATABASES, @current_to = "to '/work2/sybase/colorsbackup'", @textdb_to="to '/work2/sybase/textdbbackup'"
```

**Note** It is important that you back up the `text_db` database whenever you back up text indexes, since that database contains the metadata for all text indexes.

`sp_text_dump_database` may fail on Solaris if the required file size is greater than 2GB.

See `sp_text_dump_database` on page 156.

---

**Restoring collections and text indexes from backup**

`sp_text_load_index` system procedure restores text indexes that have been backed up with the `sp_text_dump_database` system procedure.

As Database Administrator, perform the following procedures to restore your Verity collections:

1. Restore your Adaptive Server user database and `text_db` database. This returns the source tables, metadata, and `text_events` table to a consistent and predictable state. Follow the procedures described in Chapter 27, “Backing Up and Restoring User Databases,” in the *System Administration Guide*, to restore user and `text_db` databases.
Backup and recovery for the Enhanced Full-Text Search engine

2 Run sp_text_load_index to restore the Verity collection from the most recent index dump. The procedure resets the status of all text_events table entries made since the last index dump to “unprocessed” and notifies the Enhanced Full-Text Search engine to process those events.

Example:
To restore the sample_colors_db database and all of its text indexes:

1 Restore the text_db database:

```
1> use master
2> go
1> load database text_db from '/work2/sybase/textdbbackup'
2> go
```

2 Restore the sample_colors_db database:

```
1> load database sample_colors_db from '/work2/sybase/colorsbackup'
2> go
```

3 Bring the text_db and sample_colors_db databases online:

```
1> online database text_db
2> online database sample_colors_db
3> go
```

4 Restore the text index:

```
1> use sample_colors_db
2> go
1> sp_text_load_index
2> go
```

See sp_text_load_index on page 159.
CHAPTER 7

Performance and Tuning

The Enhanced Full-Text Search engine is shipped with a default configuration. You can optimize the performance of the Enhanced Full-Text Search engine by altering the default configuration so that it better reflects the needs of your site. This chapter describes ways in which you can enhance performance.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Updating existing indexes</td>
<td>79</td>
</tr>
<tr>
<td>Increasing query performance</td>
<td>80</td>
</tr>
<tr>
<td>Reconfiguring Adaptive Server</td>
<td>81</td>
</tr>
<tr>
<td>Reconfiguring the Enhanced Full-Text Search engine</td>
<td>82</td>
</tr>
<tr>
<td>Using sp_text_notify</td>
<td>83</td>
</tr>
<tr>
<td>Configuring multiple Enhanced Full-Text Search engines</td>
<td>84</td>
</tr>
<tr>
<td>Multiple users</td>
<td>85</td>
</tr>
<tr>
<td>File Descriptors and Enhanced Full-Text Search</td>
<td>86</td>
</tr>
</tbody>
</table>

Updating existing indexes

The amount of time it takes to update records in a text index can be reduced by enabling (turning on) trace flag 11 or trace flag 12, or both:

- Enabling trace flag 11 disables Verity collection optimization. This means that Verity does not optimize the text index after you issue sp_text_notify, which is a performance gain. If trace flag 11 is turned off (the default), the Enhanced Full-Text Search engine calls Verity to optimize the text index at the end of sp_text_notify processing, which can delay the completion of sp_text_notify.

With Enhanced Full-Text Search Specialty Data Store, you can use sp_optimize_text_index to optimize a text index at a later time if trace flag 11 is enabled. See sp_optimize_text_index on page 147.
Enabling trace flag 12 disables the Enhanced Full-Text Search engine from returning sp_statistics information. If trace flag 12 is turned off (the default), an update statistics command is issued to the Enhanced Full-Text Search engine, which can delay the completion of sp_text_notify.

If updates to the text index occur as often as every few seconds, you may improve performance by disabling the update statistics processing and the Verity optimization, or both, for most of the updates.

Trace flags 11 and 12 can be enabled and disabled interactively using sp_traceon and sp_traceoff in the Enhanced Full-Text Search engine.

Increasing query performance

Two issues can significantly improve query performance:

- Limiting the number of rows returned by the Enhanced Full-Text Search engine
- Ensuring the correct join order for queries

Limiting the number of rows

Use the max_docs pseudo column to limit the number of rows returned by the Enhanced Full-Text Search engine. The fewer the number of rows returned by the Enhanced Full-Text Search engine, the faster Adaptive Server can process the join between the source table and the index table.

Ensuring the correct join order for queries

The more tables and text indexes that are listed in a join, the greater the chance that the query will run slowly because of incorrect join order. Queries run fastest when the text index is queried first during a join between the text index and one or more tables.

To ensure correct join order:

- Make sure that a unique clustered or nonclustered index is created on the IDENTITY column of the table being indexed
• Limit joins to one base table and one text index

If a query is running slowly, use showplan or enable trace flag 11205, and examine the join order. Trace flag 11205 dumps remote queries to the Adaptive Server error log file. The fastest queries contain an index_any search condition in the where clause and query the text index first.

The slowest queries contain the id column in the text index where clause and query the indexed table first. In this case, rewrite the query or use forceplan to force the join order that is listed in your query. For more information about forceplan, see Chapter 3, “Advanced Optimizing Techniques,” in the Performance and Tuning Guide: Optimizer and Abstract Plans.

Reconfiguring Adaptive Server

You can improve the performance of the Enhanced Full-Text Search engine by resetting the Adaptive Server configuration parameters discussed in this section. (For information about setting configuration parameters with sp_configure, see Chapter 4, “Setting Configuration Parameters,” in the System Administration Guide.)

cis cursor rows

cis cursor rows parameter specifies the number of rows received by Adaptive Server during a single fetch operation. The default number for cis cursor rows is 50. Increasing this number increases the number of rows received by Adaptive Server from the Enhanced Full-Text Search engine during a fetch operation. However, the number you set for cis cursor rows, the more memory Adaptive Server dynamically allocates to return the result set.
Reconfiguring the Enhanced Full-Text Search engine

**cis packet size**

cis packet size determines the number of bytes contained in a single network packet. The default for cis packet size is 512. You must specify values for this parameter in multiples of 512. Increasing this parameter improves the performance of the Enhanced Full-Text Search engine because, with a larger packet size, it returns fewer packets for each query. However, the larger the number you set for cis packet size, the more memory Adaptive Server allocates for that parameter.

The cis packet size parameter is dynamic; you do not need to restart Adaptive Server for this parameter to take effect.

---

**Note** If you change the cis packet size, you must also change max_packetsize in the Enhanced Full-Text Search engine configuration file to the same value. If Component Integration Services is used to access other remote servers, the max network packet size on those servers must be increased as well.

---

You must restart the Enhanced Full-Text Search engine for the max_packetsize parameter to take effect.

---

**Reconfiguring the Enhanced Full-Text Search engine**

You can improve the performance of the Enhanced Full-Text Search engine by reconfiguring the Enhanced Full-Text Search engine configuration parameters discussed in this section. See “Modifying the configuration parameters” on page 67.

---

**batch_size**

batch_size determines the number of rows per batch the Enhanced Full-Text Search engine indexes. batch_size has a default of 500 (that is, 500 rows of data indexed per batch). Performance improves if you increase the size of the batches that are indexed. However, the larger the batch size, the more memory the Enhanced Full-Text Search engine allocates for this parameter.

When considering how large to set batch_size, consider the size of the data on which you are creating a text index. When creating the text index, the Enhanced Full-Text Search engine allocates memory equal to (in bytes):

---
(amount of space needed for data) x (batch_size) = memory used

For example, if the data you are indexing is 10,000 bytes per row, and batch_size is set to 500, then the Enhanced Full-Text Search engine must allocate almost 5MB of memory when creating the text index.

Base the batch size you choose on the typical size of your data and the amount of memory available on your machine.

**min_sessions and max_sessions**

min_sessions and max_sessions determine the minimum and maximum number of user connections allowed for the Enhanced Full-Text Search engine. Each user connection requires approximately 5MB of memory. Do not set max_sessions to an amount that exceeds your available memory. Also, because the memory for min_sessions is allocated at start-up, if you set the number for min_sessions extremely high (to allow for a large number of user connections), a large percentage of your memory is dedicated to user connections for the Enhanced Full-Text Search engine.

You may improve the performance of the Enhanced Full-Text Search engine by setting min_sessions equal to the average number of user sessions that will be used. Doing so prevents the Enhanced Full-Text Search engine from having to allocate memory at the start of the user session.

**Using sp_text_notify**

Review the needs of your site before you decide how often to issue sp_text_notify.
Configuring multiple Enhanced Full-Text Search engines

Using `sp_text_notify` produces a load on the Enhanced Full-Text Search engine as the system procedure reads the data and updates the text collections. Depending on the size of this load, the performance hit for issuing `sp_text_notify` can be substantial. Because of the performance implications, you must determine how up-to-date the indexes need to be. If they must be close to real-time, then issue `sp_text_notify` frequently (as often as every 5 seconds). However, if your indexes do not need to be that current, you may want to wait until the system is not active before you issue `sp_text_notify`.

Note: You cannot issue `sp_text_notify` from within a transaction.

Configuring multiple Enhanced Full-Text Search engines

For tables that are used frequently, you can improve performance by placing the text indexes for these tables on separate Enhanced Full-Text Search engines. Performance improves because users can spread their queries over a number of Enhanced Full-Text Search engines, instead of sending all queries to a single engine. Each Adaptive Server can connect to multiple Enhanced Full-Text Search engines, but each Enhanced Full-Text Search engine can connect to only one Adaptive Server.

Creating multiple Enhanced Full-Text Search engines at start-up

If you are initially creating multiple Enhanced Full-Text Search engines, you can edit the `installtextserver` script so that it includes all of those Enhanced Full-Text Search engines. See “Editing the `installtextserver` script” on page 15.

Adding Enhanced Full-Text Search engines

You can add Enhanced Full-Text Search engines after your initial start-up by issuing `sp_addserver` from `isql`:

```
sp_addserver server_name [, server_class [, physical_name]]
```
where:

- `server_name` is the name used to address the server on your system (in this case, the Enhanced Full-Text Search engine).
- `server_class` identifies the category of server being added. For the Enhanced Full-Text Search engine, the value is “sds”.
- `physical_name` is the name in the interfaces file used by the server `server_name`.

For more information, see `sp_addserver` in the Reference Manual.

For example, to add a Enhanced Full-Text Search engine named BLUE, enter:

```
sp_addserver  BLUE,  sds,  BLUE
```

After you configure and start the Enhanced Full-Text Search engine, you can use the following syntax to see if you can connect to the Enhanced Full-Text Search engine from the Adaptive Server:

```
server_name...sp_show_text_online
```

For example, to connect to a server named BLUE, enter:

```
BLUE...sp_show_text_online
```

### Configuring additional Enhanced Full-Text Search engines

Each Enhanced Full-Text Search engine requires its own:

- Interfaces file entry
- Configuration file

All Enhanced Full-Text Search engines use the same database (named `text_db` by default) for storing text index metadata and the same `vesaux` and `vesauxcol` tables.

### Multiple users

The following tips may help avoid deadlocks with multiple users:

1. Make sure Adaptive Server is using the same number of connections as the Enhanced Full-Text Search. 100 is the default.
File Descriptors and Enhanced Full-Text Search

Enhanced Full-Text Search makes extensive use of file descriptors when executing searches. With concurrent searches and large text indexes, this may cause a connection to receive this error message:

```
ERRORMSG, Error (): Available files (-1) less than min 5
```

This message indicates that the Enhanced Full-Text Search process has run out of file descriptors, based on the limit for the process. If you see this error message, raise the process limit for file descriptors for the Enhanced Full-Text Search process.

`max_session_fd` limits the number of file descriptors an Enhanced Full-Text Search session can allocate. You can use this in situations where the file descriptors limit cannot be raised.

The `max_session_fd` defaults to 0, which means that each session is limited only by the Enhanced Full-Text Search process file descriptors limit. The minimum setting for `max_session_fd` is 5.

If you are seeing the error message above and you cannot raise the file descriptor limit for the Enhanced Full-Text Search, the best way to gauge how to set this parameter is to take the 

\[
\frac{(\text{max file descriptors}) - 20}{\text{(max concurrent connections)}}
\]

For example, if the file descriptors limit is set to 1024 and the maximum number of concurrent connections to the Enhanced Full-Text Search is 50:

\[
\frac{(1024) - 20}{50} \approx 20
\]

sp_configure "user connections", 100

2. Make sure the `vesaux`, `vesauxcol` and `text_events` tables (in the model, or in each of your new databases) are using row-level locking.

   - For existing tables: `alter table table_name lock datarows`
   - For new tables: `create table ... lock datarows`

3. For large batches of commands, try to break them into smaller transactions.

4. If deadlocks still occur, increase the number of locks available to Adaptive Server, and adjust the row lock promotion settings. See the `System Administration Guide`.

File Descriptors and Enhanced Full-Text Search

Enhanced Full-Text Search makes extensive use of file descriptors when executing searches. With concurrent searches and large text indexes, this may cause a connection to receive this error message:

```
ERRORMSG, Error (): Available files (-1) less than min 5
```

This message indicates that the Enhanced Full-Text Search process has run out of file descriptors, based on the limit for the process. If you see this error message, raise the process limit for file descriptors for the Enhanced Full-Text Search process.

`max_session_fd` limits the number of file descriptors an Enhanced Full-Text Search session can allocate. You can use this in situations where the file descriptors limit cannot be raised.

The `max_session_fd` defaults to 0, which means that each session is limited only by the Enhanced Full-Text Search process file descriptors limit. The minimum setting for `max_session_fd` is 5.

If you are seeing the error message above and you cannot raise the file descriptor limit for the Enhanced Full-Text Search, the best way to gauge how to set this parameter is to take the \((\text{max file descriptors}) - 20)/(\text{max concurrent connections})\).

For example, if the file descriptors limit is set to 1024 and the maximum number of concurrent connections to the Enhanced Full-Text Search is 50:

\[(1024) - 20)/50 =\approx 20\]
The more available file descriptors to an Enhanced Full-Text Search session, the better.

Because the maximum number of file descriptors that each Enhanced Full-Text Search session uses is being limited, performance may be decreased. Sybase recommends that you use `max_session_fd` carefully.

You must restart the Enhanced Full-Text Search process for `max_session_fd` changes to take effect.
CHAPTER 8

Verity Topics

This chapter is reproduced with permission from Verity. It is provided to give Enhanced Full-Text Search users insight into Verity Topics.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are topics?</td>
<td>89</td>
</tr>
<tr>
<td>Using a topic outline file</td>
<td>90</td>
</tr>
<tr>
<td>Making topics available</td>
<td>91</td>
</tr>
<tr>
<td>Knowledge bases of topics</td>
<td>91</td>
</tr>
<tr>
<td>Structure of topics</td>
<td>93</td>
</tr>
<tr>
<td>Maximum number of topics</td>
<td>96</td>
</tr>
<tr>
<td>Verity query language</td>
<td>97</td>
</tr>
<tr>
<td>Sample topic outlines</td>
<td>102</td>
</tr>
<tr>
<td>Operator reference</td>
<td>103</td>
</tr>
<tr>
<td>Modifier reference</td>
<td>113</td>
</tr>
<tr>
<td>Weights and document importance</td>
<td>115</td>
</tr>
<tr>
<td>Topic scoring and document importance</td>
<td>121</td>
</tr>
<tr>
<td>Designing topics</td>
<td>124</td>
</tr>
<tr>
<td>Preparing your topic design</td>
<td>124</td>
</tr>
<tr>
<td>Topic design strategies</td>
<td>126</td>
</tr>
<tr>
<td>Designing the initial topic</td>
<td>128</td>
</tr>
</tbody>
</table>

What are topics?

A topic is a grouping of information related to a concept, or a subject area. Topics provide a convenient means by which you can encapsulate knowledge, and make it available to end users as a shared resource. By adding topics to your Verity application, users can more easily perform searches over the subject matter which the topics represent.
Using a topic outline file

Topics are combined to form knowledge bases that represent a catalog of knowledge that users can tap into when performing searches. Knowledge bases offer users the ability to find the information they want without having to compose sophisticated queries using complex syntax.

**Topic organization**

Topics organize groups of related search criteria in a format similar to that of an outline. Operators and modifiers act as the glue that joins related groups of search criteria. You can create topics as independent units, or as units with relationships to other topics in a hierarchical structure.

**Weight assignments**

You can even give some groups of search criteria more weight than other groups of search criteria in a topic’s structure. Assigning weight to search criteria affects the importance of documents selected in a search; the closer a document is to the top of the results list, the more important, or relevant, the document is to the search criteria. A search criteria weight is a number between 0.01 and 1.00. The position of a selected document in the results list can help you determine at a glance how relevant the document is compared to the search criteria.

**Using a topic outline file**

You can compose topics by creating a topic outline file.

A topic outline file is an ASCII text file in a structured format that contains topic definitions. A topic outline file might appear as follows:

```plaintext
$Control:1
art <Accrue>
*performing-arts <Accrue>
**0.80 "ballet"
**0.50 "drama"
**0.50 'dance'
**0.80 "opera"
**0.80 "symphony"
**0.90 "chamber music"
```
**"Isaac Stern"
*film <Accrue>
**directors <Filter>
/definition="title CONTAINS Truffaut"
*visual-arts <Accrue>
literature <Accrue>
philosophy <Accrue>
language <Accrue>
history <Accrue>
$$
You can create a topic outline file with any text editor.

Making topics available

The topics you make available to users must exist within a topic set that is generated using the mktopics utility. Verity topic sets generated by mktopics can be used by any Verity application. A single topic set supports a maximum of 20,000 topic definitions, and the exact number of topics allowed for one topic set depends on the Verity query language used to define them.

Setup process

Making topics available to users is a three-step process, as outlined below.

1. Create topic definitions using a topic outline file.
2. Generate a topic set. You can create a topic set using the mktopics utility. The mktopics utility creates the topic set and can also index the topics over a specific collection.
3. Import the topic set to the Enhanced Full-Text Search engine.

Knowledge bases of topics

This section discusses the principle features of knowledge bases, and the organization format used to define topics for them.

The following aspects of topic knowledge bases are covered:
Combining topics into a knowledge base

A topic is simply a grouping of information related to a concept, or a subject area. A knowledge base is a grouping of these concepts called topics. Combining topics into a knowledge base provides users with the ability to look up concepts saved as topics in a convenient fashion.

The subject area of a topic is typically identified by the topic’s name. In the example below, the subject of the topic is *performing-arts*. This topic is composed of two structural elements, its name, *performing-arts*, and its evidence topics, *ballet, musical, dance, opera, symphony*, and *drama*.

Operators and modifiers act as the glue that joins related evidence topics. Operators represent logic to be applied to evidence topics. This logic defines the qualifications of the kinds of documents you want to find. Modifiers apply further logic to evidence topics. For example, a modifier can specify that documents containing an evidence topic not be included in the list of results.

```
performing-arts ACCRUE
  ➜ WORD ballet
  ➜ WORD drama
  ➜ STEM dance
  ➜ WORD opera
  ➜ WORD symphony
  ➜ NOT-WORD mime
```
A topic’s structure becomes more sophisticated as topics are added to it. In the next example, the topic film has been added to the structure to form what is now the top-level topic, art. In this structure, performing-arts and film are subtopics of the topic art.

Sophisticated topics are composed of top-level topics, subtopics, and evidence topics. These elements determine the related subject areas of a topic. Typically, a knowledge base consists of several top-level topics. Subtopics and evidence topics can be used by multiple top-level topics.

**Structure of topics**

The structure of topics affects how the topic is interpreted during search processing. Designing topics so that they accurately express a concept involves defining a topic structure with the components described below.
Structure of topics

Top-level topics

Top-level topics are the highest topics defined in a topic structure. Top-level topics represent the subject areas you want a Verity search agent to find. In the example below you could think of literature, philosophy, languages, history, and art as top-level subtopics that comprise the top-level topic, liberal-arts.

Subtopics

Subtopics form the levels between top-level topics and evidence topics. The name of a subtopic should identify the subject area that its subtopics or evidence topics combine to describe. For example, the subtopic visual-arts includes several related words, or evidence topics, as shown below:

94 Enhanced Full-Text Search Specialty Data Store
Evidence topics

Evidence topics are the lowest units of a topic structure. Evidence topics are strings, made up of combinations of alphanumeric characters. An evidence topic can contain up to 128 alphanumeric characters.

Topic and subtopic relationships

Each topic and its associated subtopics form a hierarchical parent and child relationship. In the example below, the subtopics performing-arts, film, visual-arts, and video are children of the art topic. The art topic itself is a child of the liberal-arts topic. The liberal-arts topic could in turn be a child of successively higher parent topics within the structure.

When you use a topic to perform a search, the subject area defined by the topic includes its subtopics, their subtopics, and so on, down to the evidence topics of the structure. Topics that are not direct descendants of the topic you use are not included in the search.
In the example above, for instance, a search using the film topic would cause the Verity search engine to find documents containing information on film, motion pictures, movies, and art films. In this example, the search would not find documents related to the performing-arts, visual-arts, or video topics since these topics are not children, of the film topic. However, if the art topic was used, the search would find documents related to all the art topic’s children, which includes performing-arts, film, visual-arts, and video.

Maximum number of topics

A single topic set representing a knowledge base can consist of as many as 20,000 topics. This includes top-level topics, subtopics, and evidence topics. Topics containing as many as 1,000 subtopics may exceed memory limitations when used in a search.

Topic naming issues

- A topic name can contain up to 128 alphanumeric characters, including hyphens and underscores.
- Topic names and evidence topics are normally case-insensitive. You can name a evidence topic using all caps, as in APPLE, initial caps, as in Apple, or all lower-case, as in apple. Case is not considered when a search is performed. Thus, if your evidence topic is entered as APPLE, the Verity search engine selects documents containing “APPLE”, “Apple”, or “apple”.

You can, however, use the CASE modifier to specify that case match the entry of a evidence topic.
CHAPTER 8  Verity Topics

Verity query language

This section describes the Verity Query Language, consisting of operators and modifiers that you can use to create topics. Operators represent logic to be applied to search elements which can be combined to create a topic. This logic defines the qualifications of the kinds of documents you want to find. Modifiers apply further logic to search elements. For example, a modifier can specify that a search element be case-sensitive.

The information in this section includes the following:

- Query language summary
- Operator precedence rules
- Sample topic outlines
- Operator reference
- Modifier reference

Query language summary

The Verity Query Language consists of operators and modifiers. Both operators and modifiers represent logic to be applied to a search element. This logic defines the qualifications a document must meet to be retrieved. Operators are classified by their type, as follows:

- Evidence operators
- Proximity operators
- Relational operators
- Concept operators
- Boolean operators

Modifiers extend the logic applied by operators and are used in combination with operators.
Evidence operators

Evidence operators expand a search word into a list of related words that are then searched for as well. When you perform a search using an evidence operator, documents containing one or more occurrences of the words in the expanded word list are documents containing the word specified, as well as its synonyms. Documents retrieved using evidence operators are not relevance-ranked unless you use the MANY modifier. See “MANY modifier” on page 114. Table 8-1 describes each evidence operator.

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORD</td>
<td>Selects documents that include one or more instances of a word you specify.</td>
</tr>
<tr>
<td>STEM</td>
<td>Selects documents that include one or more variations of the search word you specify.</td>
</tr>
<tr>
<td>THESAURUS</td>
<td>Selects documents that contain one or more synonyms of the word you specify.</td>
</tr>
<tr>
<td>WILDCARD</td>
<td>Selects documents that contain matches to a character string containing variables.</td>
</tr>
<tr>
<td>SOUNDEX</td>
<td>Selects documents that include one or more words that “sound like,” or whose letter pattern is similar to, the word specified.</td>
</tr>
<tr>
<td>NEAR/N</td>
<td>Expands the search to include the word you enter plus words that are similar to the query term. This operator performs “approximate pattern matching” to identify similar words.</td>
</tr>
</tbody>
</table>

Proximity operators

Proximity operators specify the relative location of specific words in the document; that is, specified words must be in the same phrase, paragraph, or sentence for a document to be retrieved. In the case of the NEAR and NEAR/N operators, retrieved documents are relevance-ranked based on the proximity of the specified words. When proximity operators are nested, the ones with the broadest scope should be used first; that is, phrases or individual words can appear within SENTENCE or PARAGRAPH operators, and SENTENCE operators can appear within PARAGRAPH operators. Table 8-2 describes each proximity operator.
Table 8-2: Proximity operators

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN</td>
<td>Selects documents that contain specified values in one or more document zones. A document zone represents a region of a document, such as the document’s summary, date, or body text.</td>
</tr>
<tr>
<td>PHRASE</td>
<td>Selects documents that include a phrase you specify. A phrase is a grouping of two or more words that occur in a specific order.</td>
</tr>
<tr>
<td>SENTENCE</td>
<td>Selects documents that include all of the words you specify within a sentence.</td>
</tr>
<tr>
<td>PARAGRAPH</td>
<td>Selects documents that include all of the search elements you specify within a paragraph.</td>
</tr>
<tr>
<td>NEAR</td>
<td>Selects documents containing specified search terms within close proximity to each other.</td>
</tr>
<tr>
<td>NEAR/N</td>
<td>Selects documents containing two or more words within N number of words of each other, where N is an integer.</td>
</tr>
</tbody>
</table>

Relational operators

Relational operators search document fields (such as AUTHOR) that have been defined in the collection. These operators perform a filtering function by selecting documents that contain specified field values. The fields that are used with relational operators can contain alphanumerical characters. Documents retrieved using relational operators are not relevance-ranked, and you cannot use the MANY modifier with relational operators.

When creating topics, relational operators are always used in conjunction with the special FILTER operator. See the example under the topic “visual-arts” in “Sample Topic Outlines” later in this section for the proper syntax.

A number of relational operators are available for numeric and date comparisons, including the following: = (equals), > (greater than), >= (greater than or equal to), < (less than), <= (less than or equal to).

A number of relational operators are available for text comparisons, including the following.
Table 8-3: Relational operators

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTAINS</td>
<td>Selects documents by matching the word or phrase you specify with values stored in a specific document field.</td>
</tr>
<tr>
<td>MATCHES</td>
<td>Selects documents by matching the character string you specify with values stored in a specific document field.</td>
</tr>
<tr>
<td>STARTS</td>
<td>Selects documents by matching the character string you specify with the starting characters of the values stored in a specific document field.</td>
</tr>
<tr>
<td>ENDS</td>
<td>Selects documents by matching the character string you specify with the ending characters of the values stored in a specific document field.</td>
</tr>
<tr>
<td>SUBSTRING</td>
<td>Selects documents by matching the character string you specify with a portion of the strings of the values stored in a specific document field.</td>
</tr>
</tbody>
</table>

Concept operators

Concept operators combine the meaning of search elements to identify a concept in a document. Documents retrieved using concept operators are relevance-ranked. Table 8-4 describes each concept operator.

Table 8-4: Concept operators

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>Selects documents that contain all of the search elements you specify.</td>
</tr>
<tr>
<td>OR</td>
<td>Selects documents that show evidence of at least one of your search elements.</td>
</tr>
<tr>
<td>ACCRUE</td>
<td>Selects documents that include at least one of the search elements you specify.</td>
</tr>
</tbody>
</table>

Boolean operators

Boolean operators can be assigned to topics to retrieve documents containing any or all of the children of that topic. Unlike topics created using the concept operators, Boolean operators do not accept weights. Table 8-5 describes each Boolean operator.
Table 8-5: Boolean operators

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>Selects documents that contain all children of a topic.</td>
</tr>
<tr>
<td>ANY</td>
<td>Selects documents that contain at least one of the children of a topic.</td>
</tr>
</tbody>
</table>

Modifiers

Modifiers affect the behavior of operators. Table 8-6 describes each modifier.

Table 8-6: Modifiers

<table>
<thead>
<tr>
<th>Operator name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE</td>
<td>Performs a case-sensitive search.</td>
</tr>
<tr>
<td>MANY</td>
<td>Counts the density of words or phrases in a document and produces a relevance-ranked score for the retrieved documents.</td>
</tr>
<tr>
<td>NOT</td>
<td>Excludes documents that show evidence of the specified word or phrase.</td>
</tr>
<tr>
<td>ORDER</td>
<td>Specifies the order in which search elements must occur.</td>
</tr>
</tbody>
</table>

Operator precedence rules

The Verity search engine uses precedence rules to determine how operators can be assigned. These rules state that some operators rank higher than others when assigned to topics, and affect how document selections are performed.

Table 8-7 describes how precedence rules apply to operators.
To avoid a precedence violation, do not use ANY or ALL in a parent topic whose child topic includes a concept operator (AND, OR, ACCRUE). Topics that use ANY or ALL cannot have variable weights assigned to them, so you cannot use these operators in a parent topic with any child topic that allows variable weights (such as AND, OR, ACCRUE). Topics using ANY and ALL limit evaluation to present or not present (a score of 0.00 or 1.00). If the criteria are met, the children of these topics get an automatic score of 1.00; if the criteria are not met, the children of these topics get an automatic score of 0.00; so it is not meaningful to assign these children variable weights such as 0.80.

### Table 8-7: Precedence rules

<table>
<thead>
<tr>
<th>Operator</th>
<th>Precedence</th>
<th>How precedence is determined</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>Highest precedence</td>
<td>The concept operators take the highest precedence over the other operators. Thus, subtopics of topics using these operators can be assigned any of the operators listed below under “incremental precedence” or “lowest precedence.”</td>
</tr>
<tr>
<td>OR</td>
<td>Highest precedence</td>
<td></td>
</tr>
<tr>
<td>ACCRUE</td>
<td>Highest precedence</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>Incremental precedence</td>
<td>The proximity operators refer to incremental ranges that exist within a document. Subtopics of topics using these operators can be assigned their next lowest operator in the precedence order. Thus, a phrase takes precedence over a word; a sentence takes precedence over a phrase or a word; and a paragraph takes precedence over a sentence, a phrase, or a word.</td>
</tr>
<tr>
<td>PARAGRAPH</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>SENTENCE</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>NEAR</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>NEAR/N</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>PHRASE</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>ANY</td>
<td>Incremental precedence</td>
<td></td>
</tr>
<tr>
<td>WORD</td>
<td>Lowest precedence</td>
<td>The evidence operators reside at the lowest level in a topic structure. Because evidence operators are used with words contained in documents, these operators all have the same precedence.</td>
</tr>
<tr>
<td>STEM</td>
<td>Lowest precedence</td>
<td></td>
</tr>
<tr>
<td>SOUNDEX</td>
<td>Lowest precedence</td>
<td></td>
</tr>
<tr>
<td>WILDCARD</td>
<td>Lowest precedence</td>
<td></td>
</tr>
<tr>
<td>THESAURUS</td>
<td>Lowest precedence</td>
<td></td>
</tr>
</tbody>
</table>

The following are the same topics as you would create them in a topic outline file:

```plaintext
$Control:1
art <Accrue>
```
*performing-arts <Or>
  **0.80 "drama"
  **0.50 "theater"
  **0.80 ‘dance’
*film <And>
  **0.90 "cinema"
  **0.90 "documentary"
**newsreel <Filter>
/definition="DATE  >= 05/01/96"
*film-makers <Accrue>
  **"Woody Allen"
*film-making <Paragraph>
  ***"direct"
  ***"produce"
*visual-arts <Accrue>
  **sculpture <In>
  /zonespec="title"
  **painters <Filter>
  /definition="Title MATCHES Famous Painters"
  **<Thesaurus>
  /wordtext="paint"
  literature <Accrue>
  *novels <Near>
  **0.80 "Proust"
  **0.80 "Remembrance" <Case>
  *american-novel <Sentence>
  ***"American"
  ***"novel"
history <Accrue>
  *<Wildcard>
  /wordtext="histor*"
music <Accrue>
  *jazz
  ***"bebop"
  **<Not> "fusion"
  *classical
  ***"Italian opera"
$$

Operator reference

Each operator is listed below alphabetically. Examples for many of these operators can be found in the topic outline in the previous section.
ACCRUE operator

Selects documents that include at least one of the search elements you specify. Valid search elements are two or more words or phrases. Selected documents are relevance-ranked.

The ACCRUE operator scores selected documents according to the presence of each search element in the document using a “the more, the better” approach: the more search elements found in the document, the better the document’s score. Several examples of the ACCRUE operator appear in the sample outline file in “Sample topic outlines” on page 102.

ALL operator

Selects documents that include all of the search elements you specify. Unlike the ACCRUE operator, you cannot assign weights when you use the ALL operator.

AND operator

Selects documents that contain all of the search elements you specify. Documents selected using the AND operator are relevance-ranked. The example in “Sample Topic Outlines” shows how the AND operator might be used with the topic “film.” In the example, only those documents that contain both search words and a date greater than or equal to 05/01/96 are selected and ranked according to their score.

ANY operator

Selects documents include at least one of the search elements you specify. Unlike the ACCRUE operator, you cannot assign weights when you use the ANY operator.

CONTAINS operator

Selects documents by matching the word or phrase you specify with values stored in a specific document field. When you use the CONTAINS operator, you specify the field name to search, and the word or phrase to search for.
With the CONTAINS operator, the words stored in a document field are interpreted as individual, sequential units. You may specify one or more of these units as search criteria. To specify multiple words, each word must be sequential and contiguous, and must be separated by a blank space. Use CONTAINS with the FILTER operator.

The syntax for CONTAINS is the same as that for MATCHES. See the example for MATCHES under the topic “visual arts” in “Sample topic outlines” on page 102. The example assumes that the field TITLE has been created for the collection.

The CONTAINS operator does not recognize non-alphanumeric characters. The CONTAINS operator interprets non-alphanumeric characters as a space and treats the separated values as individual units. For example, if you have defined a slash (/) as a valid character, and you enter search criteria that include this character, as in OS/2, “OS” and “2” are treated as individual units.

The CONTAINS operator does not refer to the style.lex file for the definition of which characters are included in a word.

**ENDS operator**

Selects documents by matching the character string you specify. Use ENDS with the FILTER operator. The syntax for ENDS is the same as that for MATCHES. See the example for MATCHES under the topic “visual arts” in “Sample topic outlines” on page 102. The example assumes that the field TITLE has been created for the collection.

**= (EQUALS) operator**

Selects documents whose document field values are exactly the same as the search string you specify. Use EQUALS with the FILTER operator. The syntax for EQUALS is the same as that for GREATER THAN OR EQUAL TO. See the example for GREATER THAN OR EQUAL TO under the topic “film” in “Sample topic outlines” on page 102. The example assumes that the field DATE has been created for the collection.
FILTER operator

The special FILTER operator is used in conjunction with the relational operators to do field searches. See the example under the topic “visual-arts” in “Sample topic outlines” on page 102 for the proper syntax.

> (GREATER THAN) operator

Selects documents whose document field values are greater than the search string you specify. Use GREATER THAN with the FILTER operator. The syntax for GREATER THAN is the same as that for GREATER THAN OR EQUAL TO. See the example for GREATER THAN OR EQUAL TO under the topic “film” in “Sample topic outlines” on page 102. The example assumes that the field DATE has been created for the collection.

>= (GREATER THAN OR EQUAL TO) operator

Selects documents whose document field values are greater than or equal to the search string you specify. Use GREATER THAN OR EQUAL TO with the FILTER operator. See the example under the topic “film” in “Sample topic outlines” on page 102. The example assumes that the field DATE has been created for the collection.

< (LESS THAN) operator

Selects documents whose document field values are less than the search string you specify. Use LESS THAN with the FILTER operator. The syntax for LESS THAN is the same as that for GREATER THAN OR EQUAL TO. See the example for GREATER THAN OR EQUAL TO under the topic “film” in “Sample topic outlines” on page 102. The example assumes that the field DATE has been created for the collection.
<= (LESS THAN OR EQUAL TO) operator

Selects documents whose document field values are less than or equal to the search string you specify. Use LESS THAN OR EQUAL TO with the FILTER operator. The syntax for LESS THAN OR EQUAL TO is the same as that for GREATER THAN OR EQUAL TO. See the example for GREATER THAN OR EQUAL TO under the topic “film” in “Sample topic outlines” on page 102. The example assumes that the field DATE has been created for the collection.

IN operator

Selects documents that contain specified values in one or more document zones. A document zone represents a region of a document, such as the document’s summary, date, or body text. The IN operator only works if document zones have been defined in your collections. If you use the IN operator to search collections for which zones are not defined, no documents will be selected. In addition, the zone name you specify must match the zone names defined in your collections. Consult your collection administrator to determine which zones have been defined for specific collections. The example in “Sample topic outlines” on page 102 shows how IN might be used with the word “sculpture” and the TITLE zone.

MATCHES operator

Selects documents by matching the character string you specify with values stored in a specific document field. When you use the MATCHES operator, you specify the field name to search, and the word, phrase, or number to search for.

Unlike the CONTAINS operator, the search criteria you specify with a MATCHES operator must match the field value exactly for a document to be selected. With the MATCHES operator, any occurrence of a search string that appears as a portion of a value is not selected; only values matching the entire search string are selected.

You can use question marks (?) to represent individual variable characters within a string, and asterisks (*) to match multiple variable characters within a string.
Use MATCHES with the FILTER operator. The example in “Sample topic outlines” on page 102 shows how MATCHES might be used with the phrase “famous painters” and the TITLE field. The example assumes that the field TITLE has been created for the collection.

**NEAR operator**

Selects documents containing specified search terms within close proximity to each other. Document scores are calculated based on the relative number of words between search terms. For example, if the search expression includes two words, and those words occur next to each other in a document (so that the region size is two words long), then the score assigned to that document is 1.00. Thus, the document with the smallest region containing all search terms always receives the highest score. Documents whose search terms are not within 1000 words of each other are not selected, since the search terms are probably too far apart to be meaningful within the context of the document.

The NEAR operator is similar to the other proximity operators in the sense that the search words you enter must be found within close proximity of one another. However, unlike other proximity operators, the NEAR operator calculates relative proximity and assigns scores based on its calculations.

The example in “Sample topic outlines” on page 102 shows how NEAR might be used with the topic “novels.”

**NEAR/N operator**

Selects documents containing two or more words within N number of words of each other, where N is an integer. Document scores are calculated based on the relative distance of the specified words when they are separated by N words or less. Documents containing the specified words separated by more than N words are not selected. For example, if the search expression NEAR/5 is used to find two words within five words of each other, a document that has the specified words within three words of each other is scored higher than a document that has the specified words within five words of each other.

The N variable can be an integer between 1 and 1,024, where NEAR/1 searches for two words that are next to each other. If N is 1,000 or above, you must specify its value without commas, as in NEAR/1000.
The NEAR/N operator is similar to the other proximity operators in the sense that the search words you enter must be found within a close proximity of one another. However, unlike other proximity operators, the NEAR/N operator assigns scores based on relative proximity.

**OR operator**

Selects documents that show evidence of at least one of your search elements. Documents selected using the OR operator are relevance-ranked. The example in “Sample topic outlines” on page 102 shows how you might use OR with the topic “performing-arts.”

**PARAGRAPH operator**

Selects documents that include all of the search elements you specify within a paragraph. Valid search elements are two or more words or phrases. You can specify search elements in a sequential or a random order. Documents are retrieved as long as search elements appear in the same paragraph. The example in “Sample topic outlines” on page 102 shows you how you might use PARAGRAPH with the topic “film-making.”

**PHRASE operator**

Selects documents that include a phrase you specify. A phrase is a grouping of two or more words that occur in a specific order. You must use the PHRASE operator when you enter more than one word in the evidence field. Words with the PHRASE operator are displayed in double quotes. The example in “Sample topic outlines” on page 102 shows “Woody Allen” and “Italian opera” as uses of the PHRASE operator.

**SENTENCE operator**

Selects documents that include all of the words you specify within a sentence. You can specify search elements in a sequential or a random order. Documents are retrieved as long as search elements appear in the same sentence. The example in “Sample topic outlines” on page 102 shows how you might use SENTENCE with the topic “american-novel.”
**SOUNDEX operator**

Selects documents that include one or more words that “sound like,” or whose letter pattern is similar to, the word specified. Words have to start with the same letter as the word you specify to be selected. For example, when you use SOUNDEX with “sale,” the documents selected will include words such as “sale,” “sell,” “seal,” “shell,” “soul,” and “scale.” Documents are not relevance-ranked unless the MANY modifier is used.

**STARTS operator**

Selects documents by matching the character string you specify with the starting characters of the values stored in a specific document field. Use STARTS with the FILTER operator. The syntax for STARTS is the same as that for MATCHES. See the example for MATCHES under the topic “visual arts” in “Sample topic outlines” on page 102. The example assumes that the field TITLE has been created for the collection.

**STEM operator**

Selects documents that include one or more variations of the search word you specify. Words with the STEM operator are displayed in single quotes. In the example in “Sample Topic Outlines,” the word “dance” is used with the STEM operator. Documents selected will therefore include words such as “dances,” “danced,” and “dancing,” as well as “dance.”

**SUBSTRING operator**

Selects documents by matching the character string you specify with a portion of the strings of the values stored in a specific document field. The characters that comprise the string can occur at the beginning of a field value, within a field value, or at the end of a field value. The syntax for SUBSTRING is the same as that for MATCHES. See the example for MATCHES under the topic “visual arts” in “Sample Topic Outlines.” The example assumes that the field TITLE has been created for the collection.
THESAURUS operator

Selects documents that contain one or more synonyms of the word you specify. For example, when you use the word “altitude” with the THESAURUS operator, the documents selected will include words such as “height” and “elevation.” Documents are not relevance-ranked unless the MANY modifier is used.

TYPO/N operator

The TYPO/N operator expands the search to include the specified word plus words that are similar. The optional N variable specifies the maximum number of errors between the query and the matched term. For example, TYPO mouse returns documents that include words such as “house,” “louse,” and “moose.”

WILDCARD operator

Selects documents that contain matches to a character string containing variables. The WILDCARD operator lets you define a search string with variables, which can be used to locate related word matches in documents. The example in “Sample topic outlines” on page 102 shows how you might use the string “histor*” to search for words such as “history,” “historical,” and “historian.” Documents are not relevance-ranked unless the MANY modifier is used.

Using wildcard special characters

You can use the following wildcard characters to represent variable portions of search strings with the WILDCARD operator.

<table>
<thead>
<tr>
<th>Character</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>Specifies one of any alphanumeric character, as in ?an, which locates “ran,” “pan,” “can,” and “ban.” It is not necessary to specify the WILDCARD operator when you use the question mark. The question mark is ignored in a set ([ ]) or in an alternative pattern ( { }).</td>
</tr>
</tbody>
</table>
Searching for nonalphanumeric characters

Remember that you can search for nonalphanumeric characters only if the style.lex file used to create the collections you are searching is set up to recognize the characters you want to search for. Consult your collection administrator for information.

Searching for wildcard characters as literals

The wildcard characters listed above are interpreted as wildcard characters, not literal characters, unless they are delimited by a backslash (\). If you want a wildcard character to be interpreted as a literal in a wildcard string, you must precede the character with a backslash. For example, to match the literal asterisk (*) in a wildcard string, you delimit the character as follows:

\texttt{<WILDCARD> a\^} 

Searching for special characters as literals

The following nonalphanumeric characters perform special, internal functions, and by default are not treated as literals in a wildcard string:

- comma,
- left and right parentheses ( )

<table>
<thead>
<tr>
<th>Character</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>Specifies zero or more of any alphanumeric character, as in corp*, which locates “corporate,” “corporation,” “corporal,” and “corpulent.” It is not necessary to specify the WILDCARD operator when you use the asterisk, and you should not use the asterisk to specify the first character of a wildcard string. The asterisk is ignored in a set ([ ]) or in an alternative pattern ({ }).</td>
</tr>
<tr>
<td>[ ]</td>
<td>Specifies one of any character in a set, as in &lt;WILDCARD&gt; ‘c[auo]t’, which locates “cat,” “cut,” and “cot.” Note that you must enclose the word which includes a set in backquotes (‘), and there can be no spaces in a set.</td>
</tr>
<tr>
<td>{ }</td>
<td>Specifies one of each pattern separated by a comma, as in &lt;WILDCARD&gt; ‘bank{s,er,ing}’, which locates “banks,” “banker,” and “banking.” Note that you must enclose the word which includes a pattern in backquotes (‘), and there can be no spaces in a set.</td>
</tr>
<tr>
<td>^</td>
<td>Specifies one of any character not in the set, as in &lt;WILDCARD&gt; ‘st[^oa]ck’, which excludes “stock” and “stack” but locates “stick” and “stuck.” Note that the caret (^) must be the first character after the left bracket ([) that introduces a set.</td>
</tr>
<tr>
<td>-</td>
<td>Specifies a range of characters in a set, as in &lt;WILDCARD&gt; ‘c[a-r]t’, which locates every three-letter word from “cat” to “crt.”</td>
</tr>
</tbody>
</table>
• double quotation mark “
• backslash \
• at sign @
• left curly brace {
• left bracket [
• less than sign <
• backquote `”

To interpret special characters as literals, you must surround the whole wildcard string in backquotes (‘). For example, to search for the wildcard string “a{b”, you surround the string with backquotes, as follows:

<WILDCARD> `a{b`

To search for a wildcard string that includes the literal backquote character (‘), you must use two backquotes together and surround the whole wildcard string in backquotes (‘), as follows:

<WILDCARD> `“n`t`

You can only search on backquotes if the style.lex file used to create the collections you are searching is set up to recognize the backquote character. Consult your collection administrator for information.

**WORD operator**

Selects documents that include one or more instances of a word you specify. Words with the WORD operator are displayed in double quotes. The example in “Sample topic outlines” on page 102 displays many instances of the WORD operator.

**Modifier reference**

Modifiers further specify the behavior of operators. For example, you can use the CASE modifier with an operator to specify that the case of the search word you enter be considered a search element as well. Modifiers include CASE, MANY, NOT, and ORDER, which are described below.
**CASE modifier**

Use the CASE modifier with the WORD or WILDCARD operator to perform a case-sensitive search, based on the case of the word or phrase specified.

By default, documents containing any occurrences of a search word or phrase are retrieved regardless of case. To use the CASE modifier, you simply enter the search word or phrase as you wish it to appear in retrieved documents – in all uppercase letters, in mixed uppercase and lowercase letters, or in all lowercase letters. The example in “Sample topic outlines” on page 102 shows how you might use the word “Remembrance” with the CASE modifier to refer to the first word of Proust’s novel, *Remembrance of Things Past*.

**MANY modifier**

Counts the density of words, stemmed variations, or phrases in a document, and produces a relevance-ranked score for retrieved documents. The more occurrences of a word, stem, or phrase proportional to the amount of document text, the higher the score of that document when retrieved. Because the MANY modifier considers density in proportion to document text, a longer document that contains more occurrences of a word may score lower than a shorter document that contains fewer occurrences.

The MANY modifier can be used with the following operators: WORD, WILDCARD, STEM, SOUNDEX, PHRASE, SENTENCE, PARAGRAPH and THESAURUS.

The MANY modifier cannot be used with AND, OR, ACCRUE, or relational operators.

**NOT modifier**

Use the NOT modifier with a word or phrase to exclude documents that show evidence of that word or phrase. The example in “Sample topic outlines” on page 102 shows how you might use the NOT modifier to retrieve documents that mention “bebop” but not “fusion.”
ORDER modifier

Use the ORDER modifier to express the order in which search elements must occur. If search values do not occur in the specified order in a document, the document is not selected. Always place the ORDER modifier just before the operator.

You can only use the ORDER modifier with the operators ALL, PARAGRAPH, SENTENCE, and NEAR/N.

Weights and document importance

This section describes assigning weights to search criteria in topics, and the affect of weights on selected documents. The specific information covered includes:

- Which operators accept weights
- How weights affect importance
- Assigning weights
- Topic scoring and document importance

Topic weights

When processing a search agent, the Verity search engine calculates a score for each selected document behind the scenes. A document score can be in the range from 1.0 to 0.01. The higher a document’s score, the more relevant it is. Using the score assignments for documents selected by a search agent, Verity applications can present relevance-ranked results in descending order to application users.

The ranking of documents is determined by the elements which comprise your search criteria. Document ranking can be affected depending on whether the search criteria includes topics, and whether topics include weights.
Weights and document importance

When creating topics, you can assign weights to the topic structure to indicate the relative importance of specific aspects of the topic definition. For example, you may be interested in two related subjects, but one subject is more important than another. You do not have to assign weights when you compose topics because default weights are assigned as appropriate when a topic set is indexed. However, by assigning weights you can fine-tune the importance of things you are looking for.

Which operators accept weights

Weights are used in conjunction with operators to compute scores for parent and child topics during a search. The weight you assign to a topic child can be a number between 0.01 and 1.00. A child’s weight indicates its importance relative to the other children that have been defined for its parent. The higher a child’s weight, the more important that child is considered to be with respect to its siblings.

Weights can only be assigned to the children of topics that use the concept operators, as follows:

- AND
- OR
- ACCRUE

Topics that use the proximity operators SENTENCE and PARAGRAPH, cannot be assigned a weight. These operators assume a simple “yes” or “no” presence for their children.

If a topic assigned a proximity operator is, in turn, the child of a topic which has been assigned a concept operator, such as the AND operator, that child can be assigned a weight.

It is not mandatory that you assign weights to the children of a topic just because the operator can accept weighted children. When weights are not assigned, the child has an automatic weight assignment based on its operator. Children of topics using AND and OR operators assume a weight of 1.00, and children of topics using the ACCRUE operator assume a weight of 0.50. If these operators are changed – for example, if an OR operator is changed to an ACCRUE operator – the weights of children that have not been specifically assigned a weight change accordingly. Thus, if an unweighted child of an AND topic has an assumed weight of 1.00, this assumed weight changes to 0.50 if the operator is changed to ACCRUE.
If you assign a variable weight to a topic child, then change the operator used with the parent to one that does not accept weighted children, such as the SENTENCE operator. The Verity search engine automatically assumes a weight of 1.00 while this operator is in effect. If the operator is subsequently changed to one which accepts variable-weighted children, the previously-assigned variable weights become effective once again.

How weights affect importance

When you assign a weight to the child of a topic that uses a concept operator, you specify the relative contribution of that child to the overall score produced by a topic. The higher the weight you assign to the child, the higher selected documents that contain that child appear in the list of results. Thus, weights directly affect the importance, or ranking, of selected documents.

For example, assume you have the following topic:

```
1.00 VOED 80286
1.00 VOED 80386
0.80 VOED 486
0.80 VOED 386
0.80 VOED 286
0.40 STEM clone
```

The evidence topics 80286 and 80386 (which describe the microprocessors used in PC products) have an automatic weight assignment of 1.00. The evidence topics 486, 386, and 286 have a relatively high probability of referring to their parent topic, so these evidence topics are assigned weights of 0.80. The evidence topic clone may or may not refer to PC clones at all; therefore, this evidence topic is assigned a weight of 0.40.
Weights and document importance

A search agent using this topic and its assigned weights might produce the following scores for the matched documents:

<table>
<thead>
<tr>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 01-oct-98 New Toshiba Portable Desktop Computers Offer a Serious Alternative for Desktop Users</td>
</tr>
<tr>
<td>1.00 14-Feb-98 Technology ‘Chip Set’ Unveiled for Use in Making Faster Computers</td>
</tr>
<tr>
<td>0.96 13-Feb-98 CMS Enhancements Inc. Unveils New Products</td>
</tr>
<tr>
<td>0.90 01-oct-98 Top Selling Microsoft Windows Applications Now Available in One Convenient Package</td>
</tr>
<tr>
<td>0.80 31-Oct-99 KeyCorp Successful Bidder to Acquire New York State Divisions of Empire of America Federal</td>
</tr>
<tr>
<td>0.80 15-Feb-98 Health: U.S. Birth Control Legs</td>
</tr>
</tbody>
</table>

If you change the weights of each evidence topic, the importance of your selection results are affecte, as well. In this example, if you change the weights of the evidence topic 486 to 0.60, the evidence topic 386 to 0.45, the evidence topic 286 to 0.35, and the evidence topic clone to 0.20, your selected document scores change as follows:

<table>
<thead>
<tr>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 01-oct-98 New Toshiba Portable Desktop Computers Offer a Serious Alternative for Desktop Users</td>
</tr>
<tr>
<td>1.00 14-Feb-98 Technology ‘Chip Set’ Unveiled for Use in Making Faster Computers</td>
</tr>
<tr>
<td>0.80 13-Feb-98 CMS Enhancements Inc. Unveils New Products</td>
</tr>
<tr>
<td>0.60 01-oct-98 Top Selling Microsoft Windows Applications Now Available in One Convenient Package</td>
</tr>
<tr>
<td>0.60 31-Oct-99 KeyCorp Successful Bidder to Acquire New York State Divisions of Empire of America Federal</td>
</tr>
<tr>
<td>0.50 15-Feb-98 Health: U.S. Birth Control Legs</td>
</tr>
<tr>
<td>0.45 01-oct-98 Top Selling Microsoft Windows Applications Now Available in One Convenient Package</td>
</tr>
</tbody>
</table>

Assigning weights

Assigning weights

When you assign a weight to a child, keep in mind that the weight you use reflects the importance of a child to its parent topic. The matched documents are ranked by importance to the search; thus, your selection results are directly affected by the weights you assign. If you change a weight, your selection results change as well.

Example:
The topic boeing-people includes three weighted children, binder, shrontz, and woodard, as shown below.

```
0.50 boeing-people ACCRUE
  0.80 paul binder PHRASE
  0.50 frank shrontz PHRASE
  0.60 ron woodard PHRASE
```

These subtopics are assigned various weights, as follows: the child binder is assigned a weight of 0.80, since this child is considered to be the most important of the three. The subtopic hitsman is assigned a “median” weight of 0.50, since this child is reasonably important with respect to the other two children. The subtopic johnson is assigned a low weight of 0.30, since this child is considered to be the least important with respect to the other children.

When the topic boeing-people is used for a search, the Verity search engine assumes that if the phrase “Paul Binder” is located within a document, there is a high probability that the document is relevant to a search that uses the topic boeing-people. Documents that contain the phrase “Frank Shrontz” are reasonably relevant to this search; documents that contain the phrase “Ron Woodard” are the least relevant.

Because the topic boeing-people has been assigned the ACCRUE operator, the documents displayed at the top of the results list are those that contain the greatest number of children; therefore all documents with references to all three people are given the most importance. Documents that contain only one name will be selected in an order that reflects the weights of each child. Thus, because the binder child has the highest weight, documents that include only one individual are ranked by those that refer to Paul Binder first, followed by Frank Shrontz, and finally Ron Woodard.
Weights and document importance

Automatic weight assignments
When you create a child, the Verity search engine automatically assigns a default weight of 0.50 for children of topics which use the ACCRUE operator. A weight of 1.00 is assigned automatically to children of topics that use the AND or OR operators. These default weights can be manually adjusted up or down, as described in “Changing weights” on page 120. When you create a evidence topic off of a topic that uses a proximity operator, default weight of 1.00 is assigned, and it cannot be changed.

Tips for assigning weights
When initially assigning weights, start with a weight of 0.50 for children of ACCRUE topics, and 1.00 for children of all other topics.

When assigning weights to children of topics that use the ACCRUE operator, you may select more relevant results if the children do not have overly high weights. For example, assigning all of the children of an ACCRUE topic weights of 1.00 causes all documents to have equal importance, regardless of how many of the children are present within the documents. The Verity search engine assigns equal importance to all documents containing only one child as well as for documents that contain all children, so you cannot distinguish between these documents when you view the selection results.

Assign weights between 0.80 and 0.20 for the best selection results.

Changing weights
Once you have assigned weights to children, you can test these weights by running a search using the parent topics to see if the documents you want are selected. If you find that you need to change the weights, you can edit the existing weight assigned to that subtopic or evidence topic. When you edit topic definitions in the topic outline file, you must rebuild the topic set using mktopics. For complete information about using mktopics, see your Verity application’s administration guide.
Topic scoring and document importance

When you use a topic to perform a search, the search agent starts its analysis by considering the evidence topics for that topic. If the evidence topic is present, it is given 1.00 score and is considered relevant to the search. If the evidence topic is absent, it is given a 0.00 score and is considered irrelevant to the search. If the evidence topics are weighted, the scores of the evidence topics are multiplied by the weights, then combines the resulting products in a manner specified by the operator of the parent topic. If this parent topic is, in turn, the child of another topic which is being searched, its score is multiplied by its assigned weight, and the resulting product is combined with the products of its siblings in a manner specified by the operator assigned to the parent topic. This process continues until the parent topic is reached.

The operators you use determine how parent and child scores contribute to the importance of a selected document. As each child in the topic is given an importance score, the following calculations are performed:

- If a topic uses an ACCRUE operator, the highest ranking result is taken from the products of each child’s weight and score, then adds a little to the score for each child that is present in the document.
- If a topic uses an AND operator, the products of each child’s own weight and score are compared, and the lowest product (the minimum) is taken as the score.
- If a child uses an OR operator, the products of each child’s weight and score are compared, and highest product (the maximum) is taken as the score.
- If a child uses a proximity operator (PHRASE, SENTENCE, or PARAGRAPH), or a relational operator, the child receives a score of 1.00 if the topic is present, and a score of 0.00 if the topic is not present.
- An evidence topic receives a score of 1.00 if it is present, and no score of 0.00 if it is not present.

Once the final calculations for the parent topic have been performed, a matched document becomes available to the Verity application so that users can view it with its highlights.

The following example provides a breakdown of how evidence topics and subtopics are calculated to illustrate the process by which importance is assigned to selected documents.
In the following illustration, the parent topic BOEINGCO is being used in a search.

The evidence topics of each subtopic are first checked against the documents to determine if they are present. Evidence topics that are present are assigned scores of 1.00; evidence topics that are absent are assigned a score of 0.00.

The operators at the next level of a topic structure are used to combine the scores of the evidence topics. Because the operators at this level are all proximity operators (thus, no weights assigned), they all produce scores that are either 0.00 or 1.00.

For example, assume that the following evidence topics appear within a given document:

- The evidence topic “Boeing Computer Services” appears within a phrase.
- The evidence topic “Boeing Defense” appears within a paragraph. The evidence topic “Boeing Company” appears within the document.
- The evidence topic “Ron Woodard” appears within a phrase.

The other evidence topics are only partially present, or are absent. Table 8-9 shows how the presence or absence of these evidence topics affect topic scores. The score for each topic reflects the presence of all related evidence topics, based on the operators that have been assigned to the parent topics.
Given the above topic scores, the operators at the next level of topics in the structure are calculated as follows:

- The subtopic boeing-comps, which uses the AND operator, has a score of 0.50.
- The subtopic boeing-people, which uses the ACCRUE operator, has a score of 0.50.

Finally, the topic BOEINGCO, which uses the OR operator, compares the products of each child’s weight and score, and takes the highest product (the maximum) as its score. The selected document is thus scored as 0.50.

This process is repeated for each document. The documents are sorted by the scores of the BOEINGCO topic, and displayed in ranked order.
Designing topics

This section discusses methodologies you can use to design effective topics. You can apply the methodologies and strategies described here whether you plan to compose topics using a topic outline file or one of the Verity clients. The information in this section includes the following:

- Preparing your topic design
- Topic design strategies
- Designing the initial topic

Preparing your topic design

As you prepare your topic design, consider the naming conventions you will use. Your topic names should help identify the subject matter of the kinds documents you want to find.

To ensure the best search performance, use alphanumeric characters (A through Z, and 0 through 9) for topic names. You can also use foreign characters whose ASCII value is greater than or equal to 128, as well as these symbols: $ (dollar sign), % (percentage sign), ^ (circumflex), + (plus sign), - (dash), and _ (underscore). Using other nonalphanumeric characters, may cause misinterpretation of the topic name and affect results.

Understanding your information needs

You should have an understanding of the subject areas to be addressed by your topic design and be familiar with the search requirements of users at your site. The next step is to understand your informational needs, as well as the document types to be searched.

In planning your initial topic design, keep in mind that you are developing a strategy, and the topics you define are the tactics you will use to implement that strategy.

As you develop your strategy, try to answer the following questions:

- What do you want to gain by using a Verity search agents?
- What issues are to be solved by Verity search agents?
• Who will use search agents?
• What kinds of source material will be used?
• What kinds of searches will be performed?
• How are searches currently being performed?

Consider the topics you define as questions to be asked. Just as you might ask a reference librarian at your local library for information relating to a subject area, the topics you create should pose questions when creating Verity search agents.

When considering your strategy, and how Verity search applications will be implemented to provide a solution, keep in mind that a topic you design performs several roles, as follows:
• A librarian
• A research assistant
• An information repository
• A knowledge base

Understanding your documents

To build effective topics, you must have a good understanding of the types of documents being used as information sources. For example, your documents may consist of one or more of the following types of information:
• Letters
• Memos
• Reports
• Articles

Collect representative samples of the types of documents to be searched. Note common characteristics you will need to apply to the topics you design. For example, if your documents contain important terms, acronyms, or jargon, highlight them so you can create topics that search for this text.

As you collect your document samples, identify their sources – whether they are internal sources, such as internal auditing reports; or external sources, such as e-mail messages from outside organizations. This information enables you to define the subtopics for top-level topics.
Using scanned data

If your documents are scanned into electronic files using an OCR facility, determine whether the document files will be reviewed for accuracy prior to indexing. If scanned files are reviewed, consult with reviewers to ensure that standards are applied to terms, acronyms, and jargon. If scanned files are not reviewed, note possible variations that may occur. You can develop a topic that uses an OR operator to include variations.

Categorizing document samples

Once you have collected your representative document samples and have performed an initial analysis of their contents, you may want to categorize them further. The categorization process can help you to define the top-level topics and children contained in your topic design, and help determine the operators and weights to assign.

Following are categorization examples:

- Geographic location
- Sit
- Project
- Subject area
- Date

The categorization process can help you understand the common, meaningful elements which exist in your information sources. For example, if you categorize your information by date (such as a month), it makes sense to create topics that use relational operators, such as EQUALS.

Topic design strategies

Once you have an understanding of your documents, you are ready to choose a topic design strategy. There are two topic design strategies

- The “top-down” strategy considers the major subject classifications first, followed by classifications of increasing detail.
The “bottom-up” strategy considers the detailed areas first, followed by classifications which group each detailed area by a more generalized subject.

**Top-down design**

A top-down strategy assumes you are designing a topic from the top-level topics down through the individual evidence topics of each subtopic. To design from the top down, you must adopt a taxonomy, or scientific classification approach, to creating a topic, as follows:

- Top-level topics – use general headings to identify the subject area
- Subtopics – use more specific headings to identify the primary groupings within the subject area, as well as topics which are increasingly more specific.
- Evidence topics – use important terms, acronyms, or jargon, to define the subject.

A top-down design works best when you have clearly-defined requirements. This approach is also ideal if your set of searchable documents is constantly growing or changing. With this strategy, for example, you are likely to define subjects which may not yet be evident in your information sources. Keep in mind that you can always add new topics, if you find that a number of new documents contain information which are not identified in your topic design.

If your information sources (that is, your set of indexed documents) changes constantly, specific subjects within documents may be missed, especially at the lowest levels. Periodically analyze the information being selected by your topics to ensure that topics critical to your application are current, and the appropriate information is being found.

**Bottom-up design**

A bottom-up strategy assumes you are designing a topic from the individual evidence topics up through the top-level topics that will be defined. With this strategy, your topic design objective is to select documents containing information similar to your lower-level topics.
Designing the initial topic

When you use a bottom-up design, you can start with a document which contains a good representative sample of the words or phrases you want to search for. Then you can group these words by successively higher classifications.

A bottom-up design works best when you have documents that are representative of many other documents that contain similar information. This approach is also useful when your information sources are not subject to many changes or additions.

Topic designs based on the contents of specific documents may miss related subject areas in other documents. For example, if a name is used in the sample document and that name changes in other documents, the new name may be missed in searches.

In addition, the bottom-up strategy implies that your topic design is tuned to the specific document set being used to develop your topics. These documents may not be representative of all documents contained in your information sources. Periodically review the effectiveness of your searches.

Designing the initial topic

When you have decided whether to use the top-down approach or the bottom-up approach for your initial topic design, create a topic outline to identify the topic levels to be defined.

Outlining a topic

Making a topic outline can help you determine how information will be categorized at the various levels within a topic. You can use a topic outline with the top-down or the bottom-up design approach, but it is particularly useful for the top-down approach. We recommend that every topic you build be developed as an outline first, so that you can understand the relationships between topics and subtopics, and organize them to be the most useful.

A topic outline helps you understand how information might be searched for by the people who use Verity search agents at your site. You can use a topic outline to fine-tune the information specified by topics and subtopics to pinpoint document selection. Try to do the following as you develop a topic outline:
• Identify the specific areas of information people will use when performing searches.
• Identify any related subtopics which may be grouped as children under a parent topic.
• Consider the initial level of detail to be covered by your topic design.

Keep the scope of your topic outline relatively small to begin with. A smaller, simpler topic outline is easier to define, and you can always add additional information later. As you develop your topic outline, determine how many levels your topic design will include.

Top-down topic outline example

Developing a top-down topic outline involves three steps.
• Establishing an information hierarchy
• Establishing individual search categories
• Establishing the topics to be built

As you work through these steps, meet with the people who use Verity search agents at your site to develop a topic outline that best meets their search needs, as described below.

Establishing an information hierarchy

Talk to the people at your site to learn what types of documents contain the information they need.

For example, assume you are developing a topic design for people in the medical industry to find information relating to current drug testing. Based on discussions with the people who will use Verity search agents at your site, you learn that the following types of documents are prime sources of current drug testing information:
• Research reports
• Product literature

These documents form the information sources to be searched by Verity search agents.
Designing the initial topic

Establishing individual search categories
Review the documents that will form the information sources at your site. Look for ways to categorize documents.

In our example, a review of the medical research reports and product literature shows information contained in these documents is divided into several categories. You determine that the following categories will be used to define the top-level topics in your topic design:

- Lab reports
- Clinical trials, data, or research
- Product literature

Establishing the topics to be built
Discuss categories you define with the people who create Verity search agents at your site to determine the most important concepts that selected documents should contain, and to determine the top-level topics you need to develop for each category.

For example, you determine that the category “clinical trials” includes the following top-level topics:

- product-testing
- research-methodology
Within these top-level topics, for example, the following subtopics are identified by subject-area experts:

- product-testing
- drug-names
- research-methodology
  - article-type
  - experimental-lab
  - experimental-subjects
  - organ-systems
  - age-category
  - key-aspects
  - procedural-aspects
  - study-type
  - drug-related-aspects
  - drug-administration-routes
  - geographical-areas
Once these topics are classified, you consult the people who use Verity search agents at your site to determine subtopics. Following is an example of subtopics classified as children for the topic procedural-aspects:

- product-testing → drug-names

- research-methodology

  - article-type
  - experimental-lab
  - experimental-subjects
  - organ-systems
  - age-category
  - key-aspects
  - procedural-aspects
  - study-type
  - drug-related-aspects
  - drug-administration-routes
  - geographical-areas

  - diagnosis
  - therapy
  - prevention
  - autopsy
As the topic outline is defined, you consult the people who use Verity search agents at your site to ensure the topics select meaningful documents. In the next example, a topic called drug-names enables the users at your site to search clinical trials data for drugs, based on their names.

Bottom-up topic outline example

Developing a bottom-up topic outline involves three steps.

- Identifying the subtopics that will form the lowest levels of the topic design
- Categorizing related subtopics into higher-level topics
- Establishing the top-level topic classifications

As you work through these steps, meet with the people who create Verity search agents at your site to develop a topic outline that best meets your search needs, as described below.
Identifying low-level topics

Find a document you can use as a model whose information is representative of other documents you want to find.

For example, assume you are developing a topic design to find information on the computer industry. As a start, you build a topic that searches for documents related to Apple Computer and related products.

You use the following sample as a model document for which the information is representative of other documents you want to find:

A system developed specifically for networked Apple Computer, Inc. Macintosh computers has been announced by Human Designs, Inc.

Dubbed Chorus, the floor-standing unit reportedly can contain up to 16 floating-point processors and connects to networked Macintoshes to create a multiuser desktop environment.

The product offers performance of eight million to 32 million floating-point operations per second and was designed to accommodate software development, according to the vendor. Options include an Ethernet I/O upgrade and a software simulator.

A Chorus 1 single floating-point processor entry-level system costs $9,700. A Chorus 4 configuration with four floating-point processors is available at $25,000, which includes a dedicated I/O processor with an Apple Appletalk port and system software. Both systems are upgradeable.


This document makes you decide you want to locate other documents which refer to “Appletalk” and “Macintosh,” so you define two parent topic names, apple-software and apple-hardware.

You decide you want to add additional evidence topics to select documents containing related information, such as “Macintosh.”
“Mac Classic,” “Quadra,” and “Power Mac.” In addition, you decide you want to include the evidence topics “AppleTalk,” “MacPaint,” “MacWrite,” and “MacDraw,” as related software products. You assign these evidence topics to your apple-hardware and apple-software topics, as follows:

Categorizing related subtopics

Discuss subtopics with the people who use Verity search agents at your site to determine if other subtopics exist that can be logically grouped in a category.
Designing the initial topic

In our example, some of the people who use Verity search agents are interested in finding information on personnel at Apple Computer, and others are interested in finding any documents that refer to Apple Computer. In the example below, a logical group of topics addresses several aspects of Apple Computer:

```
*apple-products*

  *apple-hardware*
  *macintosh*
  *mac classic*
  *quadra*
  *power mac*

  *appletalk*
  *macpaint*
  *macwrite*
  *macraw*

*apple-software*

*apple-company-names*

  *apple*
  *apple computer*  

*apple-misc*

  *macworld*
  *claris*  

*apple-people*

  *gil-amelio*  
```

Establishing top-level topics

Determine whether other top-level topics are necessary to find related information.
In the following example, a new topic, dec, is developed for another computer company, Digital Equipment Corporation. This topic was assigned a top-level topic and contains subtopics similar to those defined for the apple topic, as shown below.

Verity® and TOPIC® are registered trademarks of Verity, Inc.
Designing the initial topic
This appendix describes the Sybase-supplied system procedures used for updating and getting reports from system tables. Table A-1 lists the system procedures included with the Enhanced Full-Text Search engine.

### Table A-1: System procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp_check_text_index</td>
<td>Reports or fixes consistency problems in Enhanced Full-Text Search index and source tables.</td>
</tr>
<tr>
<td>sp_clean_text_events</td>
<td>Removes processed entries from the text_events table.</td>
</tr>
<tr>
<td>sp_clean_text_indexes</td>
<td>Removes text indexes that are not associated with a table.</td>
</tr>
<tr>
<td>sp_create_text_index</td>
<td>Creates an external text index.</td>
</tr>
<tr>
<td>sp_drop_text_index</td>
<td>Drops text indexes.</td>
</tr>
<tr>
<td>sp_help_text_index</td>
<td>Displays text indexes.</td>
</tr>
<tr>
<td>sp_optimize_text_index</td>
<td>Runs the Verity optimization routines.</td>
</tr>
<tr>
<td>sp_redo_text_events</td>
<td>Changes the status of entries in the text_events table and forces re-indexing of the modified table.</td>
</tr>
<tr>
<td>sp_refresh_text_index</td>
<td>Adds an entry to the text_events table reflecting updates to the corresponding source table.</td>
</tr>
<tr>
<td>sp_show_text_online</td>
<td>Displays information about databases or indexes that are currently online.</td>
</tr>
<tr>
<td>sp_text_cluster</td>
<td>Displays or modifies clustering options.</td>
</tr>
<tr>
<td>sp_text_configure</td>
<td>Displays or modifies Enhanced Full-Text Search engine configuration parameters.</td>
</tr>
<tr>
<td>sp_text_dump_database</td>
<td>Makes a backup copy of the text indexes in a database and optionally dumps the text_db and current databases.</td>
</tr>
<tr>
<td>sp_text_kill</td>
<td>Terminates all connections to a specific text index.</td>
</tr>
<tr>
<td>sp_text_load_index</td>
<td>Restores text indexes from a backup.</td>
</tr>
<tr>
<td>sp_text_notify</td>
<td>Notifies the Enhanced Full-Text Search engine that the text_events table has been modified.</td>
</tr>
<tr>
<td>sp_text_online</td>
<td>Makes a database available to Adaptive Server.</td>
</tr>
</tbody>
</table>
sp_check_text_index

Description
Reports or fixes consistency problems in the Enhanced Full-Text Search index and source tables.

Syntax
sp_check_text_index server, "index_name", "id_column", "fixit"

Parameters
server is the name of the text server.
index_name is the name of the text server.
id_column is the source identity or primary key column name.
fixit if FALSE, just reports problems. If TRUE, does not report but repairs problems.

Examples
Lists problems on the server named textsrv with the column name text.i_text:
sp_check_text_index "textsrv", "text.i_text", "id", "false"

Usage
• Before using sp_check_text_index, you must issue sp_dboption "select into", true
• This procedure:
  • Generates an sp_refresh_text_index insert for entries in the source table that do not have a matching entry in the index.
  • Generates an sp_refresh_text_index delete for entries in the index table that have no source table entry.
  • Generates an sp_refresh_text_index delete for each extra entry where duplicate index entries exist.
  • To determine the index duplicates, select all of the ID values from the index table into a temporary table. If the collection has more than 64K ID values, you must change the batch_blocksize configuration parameter from its default of 0 to 65536 to enable blocked reading of the returned Verity information. If you do not do this, Enhanced Full-Text Search attempts to real all ID values in one read and fail with a Verity error of “-27.”

Messages
None.

Permissions
Any user can execute sp_check_text_index.
sp_clean_text_events

Description
Removes processed entries from the text_events table.

Syntax
sp_clean_text_events [up_to_date]

Parameters
up_to_date
the date and time through which all processed entries will be deleted.

Examples
Removes data entered on or before January 15, 1998 at 5:00 p.m.:

```
sp_clean_text_events "01/15/98:17:00"
```

Usage
- If you do not specify the up_to_date parameter, all entries having a date less than or equal to up_to_date and whose status are set to processed is deleted.
- If you ommit up_to_date, all entries whose status is set to processed is deleted.
- Remove entries from the text_events table only after you have backed up the collection associated with the text index.
- sp_text_dump_database automatically runs.

Messages
None.

Permissions
Any user can execute sp_clean_text_events.

See also
sp_text_dump_database
**sp_clean_text_indexes**

Description: Removes indexes from the `vesaux` table that are not associated with a table.

Syntax: `sp_clean_text_indexes`

Parameters: None.

Examples: `sp_clean_text_indexes`

Usage:
- This procedure reads entries from the `vesaux` and `vesauxcol` tables, verifying that both the source table and the corresponding index table exist. If either is missing, the index is dropped.

Messages:
- Fetch resulted in an error.
- Unable to drop object definition for `index_name`!

Permissions: Any user can execute `sp_clean_text_indexes`. 
sp_create_text_index

Description
Creates a text index.

Syntax
sp_create_text_index server_name, index_table_name, table_name, "batch", column_name
[ , column_name ... ]

Parameters
server_name
is the name of the Enhanced Full-Text Search engine.

index_table_name
is the name of the index table. index_table_name has the form
[dbname.[owner.]]table, where:
- dbname is the name of the database containing the index table.
- owner is the name of the owner of the index table.
- table is the name of the index table.

table_name
is the name of the source table containing the text being indexed.
table_name has the form [dbname.[owner.]]table.

batch
the “batch” operator (must be in quotes) tells the Enhanced Full-Text Search
to reallocate every session after each batch sent to the VDK.

column_name
is the name of the column indexed by the text index.

Examples
Creates a text index and an index table named i_blurbs on the copy column of
the blurbs table:

    sp_create_text_index "blue", "i_blurbs", "blurbs", "", "copy"

Usage
- Up to 16 columns can be indexed in a single text index.
- Columns of the following datatypes can be indexed: char, varchar, nchar,
nvarchar, text, image, date, time, datetime, smalldatetime, int, smallint, and
tinyint.
- The content of option_string is not case sensitive.
- option_string uses a null string (" ") to specify “No Options.”
sp_create_text_index

- Assign the value "empty" to option_string to create a text index that you will immediately drop. This creates the Verity collection directory and the style files, but does not populate the collections. For example, when you configure an individual table for clustering, you create the text index and immediately drop it. After you edit the style.prm file, you re-create the text index. See “Editing individual style.prm files” on page 29.

- sp_create_text_index writes entries to the vesaux table and tells the Enhanced Full-Text Search engine to create the text index.

- Execution of sp_create_text_index is synchronous. The Adaptive Server process executing this system procedure remains blocked until the index is created. The time required to index large amounts of data may take as long as several hours to complete.

- When you create a text index on two or more columns, each column in the text index is placed into its own document zone. The name of the zone is the name of the column. The zones can be used to limit your search to a particular column. For more information, see “in” on page 53.

- Do not rename an index after creating.

Messages

- Cannot run sp_create_text_index from within a transaction.

- ‘column_name’ cannot be NULL.

- Column ‘column_name’ does not exist in table ‘table_name.’

- Index table mapping failed – text index creation aborted.

- Invalid text index name – ‘index_name’ already exists.

- ‘parameter’ is not in the current database.

- Server name ‘server_name’ does not exist in sysservers.

- ‘table_name’ does not exist.

- ‘table_name’ is not a valid object name.

- Table ‘table_name’ does not have an identity column – text index creation aborted.

- Text index creation failed.

- User ‘user_name’ is not a valid user in the database.

Permissions

- Any user can execute sp_create_text_index.
**sp_drop_text_index**

**Description**
Drops the index table and text indexes.

**Syntax**
```
sp_drop_text_index "table_name.index_table_name"
[,]"table_name.index_table_name"...
```

**Parameters**
- `table_name`
  - is the name of the table associated with the text indexes you are dropping.
  - `table_name` has the form `[dbname.[owner.]table]`, where:
    - `dbname` is the name of the database containing the table.
    - `owner` is the name of the owner of the table.
    - `table` is the name of the table.
  - `index_table_name` is the name of the index table and text index you are dropping.
  - `index_table_name` has the form `[dbname.[owner.]index]`

**Examples**
Drops the index table and text index associated with the `blurbs` table:
```
sp_drop_text_index "blurbs.i_blurbs"
```

**Usage**
- First, `sp_drop_text_index` issues a remote procedure call (RPC) to the Enhanced Full-Text Search engine to delete the Verity collection. Then, it removes the associated entries from the `vesaux` and `vesauxcol` tables, drops the index table, and removes the index table object definition.
  - Up to 255 indexes can be specified in a single `sp_drop_text_index` request.
  - If `database` and `owner` are not specified, the current owner and database are used.

**Messages**
- Cannot run `sp_drop_text_index` from within a transaction.
- Index `index_name` is not a text index.
- `parameter_name` is not a valid name.
- Server name `server_name` does not exist in `sys.servers`.
- Unable to drop index table `table_name`. This table must be dropped manually.
- User `user_name` is not a valid user in the `database_name` database.
- `vs_drop_index` failed with code `code_name`.

**Permissions**
Any user can execute `sp_drop_text_index`.
**sp_help_text_index**

**Description**
Displays a list of text indexes for the current database.

**Syntax**
```
sp_help_text_index [index_table_name]
```

**Parameters**
- `index_table_name` is the name of the text index you want to display.

**Examples**

**Example 1** Displays all indexes:
```
sp_help_text_index
```

**Example 2** Displays information about the text index `i_blurbs`:
```
sp_help_text_index "i_blurbs"
```

**Usage**
- `sp_help_text_index` is available only with Enhanced Full-Text Search Specialty Data Store.
- If you the `index_table_name` parameter, information about that text index is displayed. This information includes the name of the text index, the name of the Verity collection for the index, the name of the source table, the name of the IDENTITY or primary key column, and the name of the Enhanced Full-Text Search engine that created the index.
- If `index_table_name` is omitted, a list of all text indexes in the current database is displayed.

**Messages**
- No text indexes found in database `database_name`.
- Text index `index_name` does not exist in database `database_name`.
- Object must be in the current database.

**Permissions**
Any user can execute `sp_help_text_index`.
sp_optimize_text_index

Description
Performs optimization on a text index.

Syntax
sp_optimize_text_index index_table_name

Parameters
index_table_name
is the name of the text index you want to optimize. index_table_name has the form [dbname.[owner.]table], where:

- dbname is the name of the database containing the index table. If present, the owner or a placeholder is required.
- owner is the name of the owner of the index table.
- table is the name of the index table.

Examples
Optimizes the text index i_blurbs to improve query performance:

sp_optimize_text_index "i_blurbs"

Usage
- sp_optimize_text_index is available only with Enhanced Full-Text Search Specialty Data Store.
- This system procedure causes the Enhanced Full-Text Search engine to run the specified text index through the Verity optimization routines.
- sp_optimize_text_index is useful for optimizing a text index that has been updated with Verity optimization disabled (trace flag 11 turned on).
- To enable MaxClean optimization, turn on trace flag 30. This trace flag should only be used during maintenance since it could take extra time and interfere with normal usage. MaxClean is a Verity optimization feature that removes out-of-date collection files.

Messages
- 'index_table_name' is not in the current database.
- 'index_table_name' does not exist.
- Index 'index_table_name' is not a text index.
- This procedure is not supported against remote server 'server_name.'

Permissions
Any user can execute sp_optimize_text_index.

See also
“Updating existing indexes” on page 79


**sp_redo_text_events**

**Description**
Changes the status of entries in the `text_events` table and forces the reindexing of the modified columns.

**Syntax**

```
sp_redo_text_events [from_date [, to_date]]
```

**Parameters**

- **from_date**
  is the starting date and time in a date range of entries to be modified.

- **to_date**
  is the ending date and time in the specified date range of the entries to be modified.

**Examples**

Re-indexes columns that were modified between January 5, 1998 at 5:00 p.m. and February 12, 1998 at 8:30 a.m.:

```
sp_redo_text_events "01/05/98:17:00", "02/12/98:08:30"
```

**Usage**

- Resets the status to “unprocessed” for all entries in the `text_events` table that currently have a status of “processed.” The Enhanced Full-Text Search engine is notified that a reindex operation is required.

- Useful for synchronizing a text index after a recovery of the Verity collection from a backup. This procedure is run automatically during `sp_text_load_index`.

- If you omit `to_date`, all entries between `from_date` and the current date with a status of “processed” are reset to “unprocessed.”

- If you omit both `from_date` and `to_date`, all entries in the `text_events` table with a status of “processed” are reset to “unprocessed.”

**Messages**

- `to_date` cannot be specified without `from_date`.

- You have not specified the full range.

**Permissions**

Any user can execute `sp_redo_text_events`. 
sp_refresh_text_index

Description
Records modifications in the text_events table when you change the text index’s source table data.

Syntax
sp_refresh_text_index table_name, column_name, rowid, mod_type

Parameters
- table_name is the name of the source table being updated. table_name has the form [dbname.[owner.]]table, where:
  - dbname is the name of the database containing the table.
  - owner is the name of the owner of the table.
  - table is the name of the table.
- column_name is the name of the column being updated.
- rowid is the IDENTITY or primary key column value of the changed row.
- mod_type specifies the type of the change. Must be insert, update, or delete.

Examples
Records in the text_events table that you have updated the copy column of the blurbs table. The row you have updated has an id of 2.000000:

```
sp_refresh_text_index "blurbs", "copy", 2.000000, "update"
```

Usage
- The user maintains the consistency of the text index. You must run sp_refresh_text_index anytime you update source data that has been indexed so that the text_events table reflects the change. This keeps the collections in sync with your source data. The collections are not updated until you run sp_text_notify.
- You can create triggers that issue sp_refresh_text_index for non-text and non-image columns. See “Propagating changes to the text index” on page 23.

Messages
- Column ‘column_name’ does not exist in table ‘table_name.’
- Invalid mod_type specified (‘mod_type’). Correct values: INSERT, UPDATE, DELETE.
- Owner ‘owner_name’ does not exist.
- Table ‘table_name’ does not exist.
- ‘table_name’ is not a valid name.
sp_refresh_text_index

- Text event table not found.

Permissions
Any user can execute sp_refresh_text_index.

See also
sp_text_notify
**sp_show_text_online**

**Description**
Displays information about databases or text indexes that are currently online.

**Syntax**
```
sp_show_text_online server_name [,(INDEXES | DATABASES) ]
```

**Parameters**
- `server_name` is the name of the Enhanced Full-Text Search engine to which the request is sent.
- `INDEXES | DATABASES` specifies whether the request should contain data about online indexes or online databases. The default is `INDEXES`.

**Examples**
1. **Example 1** Displays all indexes that are currently online in the KRAZYKAT Enhanced Full-Text Search engine:
   ```
   exec sp_show_text_online KRAZYKAT
   ```
2. **Example 2** Displays all databases that are currently online in the KRAZYKAT Enhanced Full-Text Search engine:
   ```
   exec sp_show_text_online KRAZYKAT, DATABASES
   ```

**Usage**
- `sp_show_text_online` issues a remote procedure call to the Enhanced Full-Text Search engine to retrieve information about the indexes or the databases that are currently online.
- If the results of this procedure do not list a database, use `sp_text_online` to bring the desired database online.

**Messages**
- `sp_show_text_online` failed for server `server_name`.
- The parameter value `value` is invalid.
- The RPC sent to the server returned a failure return code.
- The second parameter must be `INDEXES` or `DATABASE`.

**Permissions**
Any user can execute `sp_show_text_online`.

**See also**
`sp_text_online`
**sp_text_cluster**

*Description* Displays or changes clustering parameters for the active thread.

*Syntax* `sp_text_cluster server_name, cluster_parameter[, cluster_value]`

*Parameters*

- **server_name** is the name of the Enhanced Full-Text Search engine.
- **cluster_parameter** is the name of the clustering parameter. Values are shown in Table A-2.
- **cluster_value** is the value you assign to the clustering parameter for the active thread. Values are shown in Table A-2.

### Table A-2: Clustering configuration parameters

<table>
<thead>
<tr>
<th>cluster_parameter</th>
<th>Values for cluster_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster_style</td>
<td>Specifies the type of clustering to use. Valid values are:</td>
</tr>
<tr>
<td></td>
<td>• fixed – generates a fixed number of clusters. The number is set by the cluster_max parameter.</td>
</tr>
<tr>
<td></td>
<td>• coarse – automatically determines the number of clusters to generate, based on fewer, coarse grained clusters.</td>
</tr>
<tr>
<td></td>
<td>• medium – automatically determines the number of clusters to generate, based on medium-sized clusters.</td>
</tr>
<tr>
<td></td>
<td>• fine – automatically determines the number of clusters to generate, based on smaller, finer-grained clusters.</td>
</tr>
<tr>
<td>cluster_max</td>
<td>Specifies the maximum number of clusters to generate when cluster_style is set to fixed. A value of 0 means that the search engine determines the number of clusters to generate.</td>
</tr>
<tr>
<td>cluster_effort</td>
<td>Specifies the amount of effort (time) that the search engine should expend on finding a good clustering. Valid values are:</td>
</tr>
<tr>
<td></td>
<td>• effort_default – the search engine spends the default amount of time. You can also use the Verity term “default” if you enclose it in double quotes (“ ”).</td>
</tr>
<tr>
<td></td>
<td>• high – the search engine spends the longest time.</td>
</tr>
<tr>
<td></td>
<td>• medium – the search engine spends less time.</td>
</tr>
<tr>
<td></td>
<td>• low – the search engine spends the least amount of time.</td>
</tr>
<tr>
<td>cluster_order</td>
<td>Specifies the order in which to return the rows within the clusters. Valid values are:</td>
</tr>
<tr>
<td></td>
<td>• “0” – indicates rows are returned in order of similarity to the cluster center. This means the first row returned for a cluster is the one that is most prototypical of the rows in the cluster.</td>
</tr>
</tbody>
</table>
|                   | • “1” – indicates that rows are returned in the same relative order in which they were submitted for clustering. For example, if cluster 1 contains the first, third and seventh rows found for the query, they will be returned in that relative order within the cluster.
Examples

**Example 1** Changes the *cluster_order* parameter to 1 for the active thread:

```
sp_text_cluster KRAZYKAT, cluster_order, "1"
```

**Example 2** Displays the current value of the *cluster_style* parameter:

```
sp_text_cluster KRAZYKAT, cluster_style
```

Usage

- The Verity clustering algorithm attempts to group similar rows together, based on the values of the clustering parameters.
- If you specify the *cluster_parameter* parameter, but omit the *cluster_value* parameter, *sp_text_cluster* displays the value of the clustering parameter that is specified.
- *sp_text_cluster* does not modify the value of the clustering configuration parameter. The *cluster_value* is valid only for the thread that is currently executing. To modify the default values, use *sp_text_configure*.
- For information on how to request a clustered result set, see “Using pseudo columns to request clustered result sets” on page 48.

Messages

- This procedure is not supported against remote server *server_name*.
- The parameter value ‘*value*’ is invalid.
- *sp_text_cluster* failed (status = *status*).

Permissions

Any user can execute *sp_text_cluster*.

See also

*sp_text_configure*
sp_text_configure

Description  Displays or changes Enhanced Full-Text Search engine configuration parameters.

Syntax

\texttt{sp\_text\_configure server\_name \[, \textit{config\_name} \[, \textit{config\_value}\] \]]}

Parameters

- \textit{server\_name} is the name of the Enhanced Full-Text Search engine.
- \textit{config\_name} is the name of the configuration parameter to be displayed or modified.
- \textit{config\_value} is the value you assign to the configuration parameter.

Examples

- **Example 1** Changes the backup destination directory to /data/backup:
  \texttt{sp\_text\_configure KRAZYCAT, backdir, "/data/backup"}

- **Example 2** Displays the backup destination directory:
  \texttt{sp\_text\_configure KRAZYCAT, backdir}

Usage

- When you execute \texttt{sp\_text\_configure} to modify a dynamic parameter:
  - The configuration and run values are updated.
  - The configuration file is updated.
  - The change takes effect immediately.

- When you execute \texttt{sp\_text\_configure} to modify a static parameter:
  - The configuration value is updated.
  - The configuration file is updated.
  - The change takes effect only when you restart the Enhanced Full-Text Search engine.

- When issued with no parameters, \texttt{sp\_text\_configure} displays a report of all Enhanced Full-Text Search engine configuration parameters and their current values.

- If you specify the \textit{config\_name} parameter, but omit the \textit{config\_value} parameter, \texttt{sp\_text\_configure} displays the report for the configuration parameter specified.

- For information on the individual configuration parameters, see “Modifying the configuration parameters” on page 67.

Messages

- Configuration value cannot be specified without a configuration option.
• This procedure is not supported against remote server ‘server_name.’
• sp_text_configure failed – possible invalid configuration option ‘config_name.’

Permissions
Any user can execute sp_text_configure.
**sp_text_dump_database**

**Description**
Makes a backup copy of a text index.

**Syntax**
```
sp_text_dump_database backupdbs [ current_to ] [,
    current_with ] [ current_stripe01 ] [ ... [,
    current_stripe31 ] ] [ textdb_to ] [ textdb_with ] [,
    textdb_stripe01 ] [ ... [ textdb_stripe31 ] ]
```

**Parameters**
- **backupdbs**
  Specifies whether the current database and the text_db database are backed up before the text index is backed up. Valid values are:
  - **CURRENT_DB_AND_INDEXES** – indicates that the current database is backed up before the text indexes are backed up.
  - **CURRENT_DB_AND_CURRENT_INDEXES** – indicates that the current database is backed up before the text indexes are backed up, and only the indexes associated with the current database are dumped.
  - **TEXT_DB_AND_INDEXES** – indicates that the text_db database is backed up before the text indexes are backed up.
  - **INDEXES_AND_DATABASES** – indicates that the current and text_db databases are backed up before the text indexes are backed up.
  - **ONLY_INDEXES** – indicates that only the text indexes are backed up.

- **current_to**
  Is the to clause of the dump database command for dumping the current database. Use this only if you specify **CURRENT_DB_AND_INDEXES** or **INDEXES_AND_DATABASES** for the `backupdbs` parameter.

- **current_with**
  Is the with clause of the dump database command for dumping the current database. Use this only if you specify **CURRENT_DB_AND_INDEXES** or **INDEXES_AND_DATABASES** for the `backupdbs` parameter.

- **current_stripe**
  Is the stripe clause of the dump database command for dumping the current database. Use this only if you specify **CURRENT_DB_AND_INDEXES** or **INDEXES_AND_DATABASES** for the `backupdbs` parameter.

- **textdb_to**
  Is the to clause of the dump database command for dumping the text_db database. Use this only if you specify **INDEXES_AND_DATABASES** for the `backupdbs` parameter. Use this only if you specify **TEXT_DB_AND_INDEXES** or **INDEXES_AND_DATABASES** for the `backupdbs` parameter.
**APPENDIX A  System Procedures**

`textdb_with`

is the with clause of the `dump database` command for dumping the `text_db` database. Use this only if you specify `TEXT_DB_AND_INDEXES` or `INDEXES_AND_DATABASES` for the `backupdbs` parameter.

`textdb_stripe`

is the stripe clause of the `dump database` command for dumping the `text_db` database. Use this only if you specify `TEXT_DB_AND_INDEXES` or `INDEXES_AND_DATABASES` for the `backupdbs` parameter.

**Examples**

**Example 1** Only text indexes are backed up:

```
sp_text_dump_database ONLY_INDEXES
```

**Example 2** The current database is dumped to `/data/db1backup` before the text indexes are backed up:

```
sp_text_dump_database CURRENT_DB_AND_INDEXES, "to '/data/db1backup'"
```

**Example 3** The `text_db` database is dumped to `/data/textdbbackup` before the text indexes are backed up:

```
sp_text_dump_database @backupdbs = "TEXT_DB_AND_INDEXES",
@textdb_to = "to '/data/textdbbackup'"
```

**Example 4** The current database is dumped to `/data/db1backup` and the `text_db` database is dumped to `/data/textdbbackup` before the text indexes are backed up:

```
sp_text_dump_database @backupdbs = "INDEXES_AND_DATABASES",
@current_to = "to '/data/db1backup'",
@textdb_to = "to '/data/textdbbackup'"
```

**Usage**

- The Enhanced Full-Text Search engine concatenates the values of `current_to`, `current_with`, and `current_stripe01` to `current_stripe31` to the string `dump database currentdbname` and then executes the `dump database` command. The output from the execution of the `dump database` command is sent to the Enhanced Full-Text Search error log.

- The Enhanced Full-Text Search engine concatenates the values of `textdb_to`, `textdb_with`, and `textdb_stripe01` to `textdb_stripe31` to the string “dump database currentdbname” and then executes the `dump database` command. The output from the execution of the `dump database` command is sent to the Enhanced Full-Text Search error log.

- All entries in the `text_events` table that have a “processed” status in the current database are deleted when all indexes have been backed up.

- The backup files for the Verity collections are stored in the directory specified in the `backDir` configuration parameter.
See references to the configuration parameter `backCmd` for customizing backups.

- The parameter value ‘value’ is invalid.
- Server name ‘server’ does not exist in sysservers.
- Attempt to dump database ‘database_name’ failed – use the dump database command.
- Attempt to backup text indexes on server ‘server_name’ failed.
- Attempt to clean text_events in database ‘database_name’ failed (date = ‘date’).
- Parameter ‘parameter_name’ is required when dumping database ‘database_name’.
- Dumping database ‘database_name’ – check Full Text Search SDS error log for status.

Any user can execute `sp_text_dump_database`.

See also `dump_database` in the Reference Manual.
**sp_text_kill**

**Description**
Terminates all connections to a specific text index.

**Syntax**
```
sp_text_kill index_table_name
```

**Parameters**
- `index_table_name` is the name of the text index from which all connections are terminated.
  - `index_table_name` has the form `[dbname.[owner.]]table`, where:
    - `dbname` is the name of the database containing the index table. If present, the `owner` or a placeholder is required.
    - `owner` is the name of the owner of the index table.
    - `table` is the name of the index table.

**Examples**
Terminates all existing connections to the text index `i_blurbs`:
```
sp_text_kill "i_blurbs"
```

**Usage**
- This system procedure causes the Enhanced Full-Text Search engine to terminate all connections to the specified index, except for the connection that initiated the request.
- Attempts to drop a text index that is currently in use will fail. `sp_text_kill` can be used to terminate all existing connections so that the index can be successfully dropped.

**Messages**
- Index ‘`index_table_name`’ is not a text index.
- This procedure is not supported against remote server ‘`server_name`.’
- ‘`index_table_name`’ does not exist.
- Only the System Administrator (SA) may execute this procedure.

**Permissions**
Only user “sa” can execute `sp_text_kill`.

**See also**
- `sp_drop_text_index`

---

**sp_text_load_index**

**Description**
Restores a text index backup.

**Syntax**
```
sp_text_load_index
```

**Parameters**
None.
sp_text_load_index

Examples

```
sp_text_load_index
```
Restores all text indexes in the current database.

Usage

- Run `sp_text_load_index` after the `text_db` database and the current database have been fully recovered.
- `sp_text_load_index` restores the Verity collections from the most recent backup. The Enhanced Full-Text Search engine then runs `sp_redo_text_events` and `sp_text_notify` to reapply all entries in the `text_events` table since the date and time the index was backed up.

Messages

- Server name ‘`server_name`’ does not exist in `sysservers`.
- Unable to restore text indexes for server ‘`server_name`’.
- This procedure is not supported against remote server ‘`server_name`’.
- Update to `text_events` table in database `database_name` failed for server ‘`server_name`’ – `text_events` not rolled forward.

Permissions

Any user can execute `sp_text_load_index`.

See also

`sp_redo_text_events`; `sp_text_notify`
sp_text_notify

Description
Notifies the Enhanced Full-Text Search engine that the text_events table has been modified.

Syntax
sp_text_notify [{true | false}] [ , server_name ]

Parameters
true
causes the procedure to run synchronously.
false
causes the procedure to run asynchronously.
server_name
is the name of the Enhanced Full-Text Search engine you are notifying.

Examples
sp_text_notify true

Usage
• You must run sp_text_notify after you issue sp_refresh_text_index to inform the Enhanced Full-Text Search engine that the source tables have been modified.
• If you do not specify true or false, sp_text_notify runs synchronously.
• If no server name is specified, all Enhanced Full-Text Search engines are notified.

Messages
• Can’t run sp_text_notify from within a transaction.
• Notification failed, server = ‘server_name.’
• Server name ‘server_name’ does not exist in sysservers.
• The parameter value ‘value’ is invalid.

Permissions
Any user can execute sp_text_notify.

See also
sp_refresh_text_index
sp_text_online

Description
Makes a database available for full-text searches to Adaptive Server.

Syntax
sp_text_online [server_name], [database_name]

Parameters
server_name
is the name of the Enhanced Full-Text Search engine.

database_name
is the name of the database that you are bringing online.

Examples
Makes the pubs2 database available for full-text searches using the Enhanced Full-Text Search engine:

sp_text_online @database_name = pubs2

Usage
• If a database is not specified, all databases are brought online for full-text searches.

• If a server name is not specified, all Enhanced Full-Text Search engines listed in the vesaux table are notified.

• With the Enhanced Full-Text Search engine, databases are brought online automatically if the auto_online configuration parameter is set to 1.

Messages
• All Databases using text indexes are now online

• Databases containing text indexes on server ‘database_names’ are now online.

• Server name server_name is now online.

• Server name ‘server_name’ does not exist in sysservers.

• The parameter value ‘value’ is invalid.

• The specified database does not exist.

• vs_online failed for server ‘server_name’.

Permissions
Any user can execute sp_text_online.
This appendix contains the following:

- The text of the default configuration file (textsvr.cfg)
- An overview of the sample_text_main.sql sample script
- A list of all the sample files provided by the Enhanced Full-Text Search engine
- An overview of the getsend program

Default textsvr.cfg configuration file

;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
; @(#) File: textsvr.cfg  1.17  07/26/99
;
; Full Text Search Specialty Data Store
; Sample Configuration File
;
; The installation procedure places this file in the
; "SYBASE" directory.
;
; Lines with a semi-colon in column 1 are comment lines.
;
; Modification History:
; ---------------------
; 11-21-97  Create file for Full Text Search SDS
; 03-02-98  Add trace flags and config values for
;          Enhanced Full Text Search SDS
; 05-26-99  remove references to sds/text
; 07-09-99  added batch block size
; 08-24-99  remove version string and correct copyright
;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;
; copyright (c) 1997, 1999
Default textsrv.cfg configuration file

`; Sybase, Inc. Emeryville, CA
; All rights reserved.
;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;
; DIRECTIONS
;
; Modifying the textsrv.cfg file:
; -------------------------------
; An installation can run the Text Search SDS product
; as supplied, with no modifications to configuration
; parameters. Default values from the executable program
; are in effect.
;
; The "textsrv.cfg" file is supplied with all configuration
; parameters commented out.
;
; The hierarchy for setting configuration values is:
;
; default value internal to the executable program (lowest)
; configuration file value (overrides default value)
; command line argument (overrides default value and *.cfg file)
;
; Command line arguments are available to override
; settings for these options:
;
; -i<file specification for interfaces file>
; -l<file specification for log file>
; -t (no arg) directs text server to write start-up
; information to stderr (default is DO NOT write start-up information)
;
; To set configuration file parameters, follow these steps:
;
; (1) If changing the server name to other than "textsrv":
; (1A) Copy "textsrv.cfg" to "your_server_name.cfg"
;      Example: text_server.cfg
; (1B) Modify the [textsrv] line to [your_server_name]
;      Example: [text_server]
;      The maximum length of "your_server_name" is 30 characters.
;
; (2) Set any configuration values in the CONFIG VALUES SECTION below.
; Remove the semi-colon from column 1.
;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;
DEFINITIONS OF TRACE FLAG AND SORT ORDER VALUES

"traceflags" parameter, for text server
Available "traceflags" values: 1,2,3,4,5,6,7,8,9,10,11,12,13
1 trace connect/disconnect/attention events
2 trace language events
3 trace rpc events
4 trace cursor events
5 log error messages returned to the client
6 trace information about indexes
7 trace senddone packets
8 write text server/Verity api interface records to the log
9 trace sql parser
10 trace Verity processing
11 disable Verity collection optimization
12 disable returning of sp_statistics information
13 trace backup operations (Enhanced Full Text Search only)

"srv_traceflags" parameter, for Open Server component of text server
Available "srv_traceflags" values: 1,2,3,4,5,6,7,8
1 trace TDS headers
2 trace TDS data
3 trace attention events
4 trace message queues
5 trace TDS tokens
6 trace open server events
7 trace deferred event queue
8 trace network requests

"sort_order" parameter
Available "sort_order" values: 0,1,2,3
0 order by score, descending (default)
1 order by score, ascending
2 order by timestamp, descending
3 order by timestamp, ascending

CONFIG VALUES SECTION

The "textsvr.cfg" file is supplied with the values commented out.
To override value(s) in the executable program:
- Set required value(s) below
- Remove the semicolon from column 1
[textsvr]
;min_sessions = 10
;max_sessions = 100
;batch_size = 500
;sort_order = 0
;defaultDb = text_db
;errorLog = textsvr.log
;language = english
;charset = iso_1
;vdkLanguage =
;vdkCharset = 850
;traceflags = 0
;srv_traceflags = 0
;max_indexes = 126
;max_packetsize = 2048
;max_stacksize = 34816
;max_threads = 50
;collDir = <txtsvr directory tree location on UNIX>/collections
;collDir = <txtsvr directory tree location on Win-NT>/collections
;vdkHome = <txtsvr directory tree location on UNIX>/verity
;vdkHome = <txtsvr directory tree location on Win-NT>/verity
;interfaces = <$SYBASE location on UNIX>/interfaces
;interfaces = <%SYBASE% location on Win-NT>/ini\sql.ini

;auto_online = 0
;backDir = <txtsvr directory tree location on UNIX>/backup
;backDir = <txtsvr directory tree location on Win-NT>/backup
;restoreCmd =
;knowledge_base =
;nocase = 0
;cluster_max = 0
;cluster_order = 0
;cluster_style = Fixed
;cluster_effort = Default
;batch_blocksize = 0
;max_session_fd = 0
The *sample_text_main.sql* script

The Enhanced Full-Text Search engine installation copies the *sample_text_main.sql* script to the `$SYBASE/$SYBASE_FTS/sample/scripts` directory. This script illustrates the following operations:

- Setting up a text index.
- Modifying data and propagating changes to the collections. This includes inserts, updates, and deletes.
- Dropping a text index.

Execution of this script is not required for installation or configuration; Sybase supplies the script as a sample.

Before you run the *sample_text_main.sql* script:

- Your Adaptive Server and Enhanced Full-Text Search engine must be configured and running.
- Use a text editor to edit the *sample_text_main.sql* script. Change “YOUR_TEXT_SERVER” to the name of your Enhanced Full-Text Search engine in Step 4 in the *sample_text_main.sql* script.
- Verify that your model database contains a text_events table. If your model database is not configured this way, you must:
  - Modify the *sample_text_main.sql* script to exit after creating the database
  - Apply the *installevent* script to the new database. See “Running the *installevent* script” on page 17.
  - Execute the remainder of the sample script

Direct the script as input to your Adaptive Server. For example, to run the *sample_text_main.sql* script on an Adaptive Server named MYSVR:

```
isql -Ulogin -Ppassword -SMYSVR -i
$SYBASE/$SYBASE_FTS/sample/scripts/sample_text_main.sql -omain.out
```

When you finish with this sample environment, log in to your Adaptive Server and drop the sample database. For example:

```
1> use master
2> go
1> drop database sample_colors_db
2> go
```
Sample files illustrating Enhanced Full-Text Search engine features

You can re-run the `sample_text_main.sql` necessary.

**Sample files illustrating Enhanced Full-Text Search engine features**

The Enhanced Full-Text Search engine supplies a set of sample files for illustrating text server operations. The files are located in the `$SYBASE/$SYBASE_FTS/sample/scripts` directory. Execution of the sample files is not required for installation, configuration, or operation of a Enhanced Full-Text Search engine.

**Custom thesaurus**

The following files illustrate how to set up and use a custom thesaurus:

- `sample_text_thesaurus.ctl` – is a sample control file.
- `sample_text_thesaurus.sql` – provides sample queries using the custom thesaurus created by the sample control file.

**Topics**

The following files illustrate how to set up and use topics:

- `sample_text_topics.otl` – is a sample outline file.
- `sample_text_topics.kbm` – is a sample knowledge base map.
- `sample_text_topics.sql` – provides sample queries using the defined topics.

**Clustering, summarization, and query-by-example**

The following files illustrate how to set up and use clustering, summarization and query-by-example:

- `sample_text_setup.sql` – creates a sample environment.
- `sample_text_queries.sql` – issues queries against the environment and drops the environment.
getsend sample program

The Enhanced Full-Text Search engine includes a program named getsend to load text or image data from a file into a column defined in Adaptive Server. The required source and header files, a makefile, and directions for building and running the program are included in the directory $SYBASE/$SYBASE_FTS/sample/source.

See the README.TXT file and getsend.c file for information on how to use the program.
getsend sample program
APPENDIX C

Unicode Support

The Unicode standard, a subset of the International Standards Organization’s ISO 10646 standard, is an international character set. Unicode is identical to the Basic Multilingual Plane (BMP) of ISO 10646, which supports all the major scripts and languages in the world. Therefore, it is a superset of all existing character sets.

Unicode:

- Provides single-source development. This means you develop an application once and it can then be localized for multiple locales and in multiple languages. By using a single unified character set, you do not have to modify your applications to take into account differences between character sets, thus reducing development, testing, and support costs.

- Allows you to mix different languages in the same database. An all-Unicode system does not require that you design your database to keep track of the character set of your data.

The Enhanced Full-Text Search engine supports Unicode. To use this feature, set the charset configuration value to utf8. This contains everything you need to set up a Unicode-enabled client/server database system.

To configure the Enhanced Full-Text Search engine to store data in Unicode format, set the charset configuration value to utf8. See “Modifying the configuration parameters” on page 67.

Note If you issue wildcard searches against data in Unicode format, turn on trace flag 15. See “Setting trace flags” on page 73.
APPENDIX D

Working with XML data

This appendix contains the following:

- The correct format for using XML data in fields and zones
- A sample of XML data indexed into text indexes with fields and zones

Correctly formatting XML data for fields and zones

You can put XML data into a text column in an Adaptive Server database, and then create a text index on that data. This requires using select statements with special syntax, which includes the Verity operators.

In the following select statement, for example, `<in>` is a Verity operator.

```sql
select t1.id
from
ti_address_tbl t1,
address_tbl t2
where
t1.id=t2.id and
t1.index_any = 'USA <in> address'
```

Sybase supports indexing XML data into text indexes with fields or zones when the XML document is well-formed and has the following format:

```xml
<address>
  <street>123 Main St.</street>
  <city>Anywhere</city>
  <state>CA</state>
  <country>USA</country>
</address>
```

This format inserts the text "123 Main St. Anywhere CA USA" into the address zone, "123 Main St." into the street zone, and so forth.
Sample XML indexing

This section provides an example of using the format in “Correctly formatting XML data for fields and zones.” This sample accomplishes the following steps:

1. Creates an empty text index on an empty table.
2. Drops the text index. This step drops the proxy tables associated with the text index in Adaptive Server, but leaves all default directories and files in the collections directory structure for EFTS text indexes.
3. Modifies the style.dft file to set the universal filter.
4. Inserts data into the table.
5. Recreates the text index.
6. Issues select requests that isolate rows based on the XML data, using the text index.

Sample in sections

Section 1

This section demonstrates creating an empty text index on an empty table.

```sql
isql -Sase1501sunbox -Usa -P -w2048 Dxml_db1
create table address_tbl
(
    id numeric(5,0)
    identity,
    xmlcol text
)
go
create unique index uidx on address_tbl(id)
go
sp_create_text_index 'textsvr', 'ti_address_tbl', 'address_tbl', "empty ", 'xmlcol'
go
(return status = 0)
quit
```
Section 2

This section demonstrates dropping an empty text index. The example drops the proxy tables associated with the Adaptive Server text index, but leaves all the default directories and files in the collections directory structure of EFTS text indexes.

```sql
isql -Sase1501sunbox -Usa -P -w2048 Dxml_db1
sp_drop_text_index 'address_tbl.ti_address_tbl'
go
(return status = 0)
quit
```

Section 3

This section demonstrates modifying the style.dft file to set the universal filter.

```none
/sy/ase1501sunbox/EFTS-15_0/collections
xml_db.dbo.ti_address_tbl/style
(uid=syuid)_sunbox> pwd
/sy/ase1501sunbox/EFTS-15_0/collections
xml_db.dbo.ti_address_tbl/style
(uid=syuid)_sunbox> ls -al
Total 74
```

```none
drwxr-xr-x 2 sybase sybase 512 Jan  8 16:14 .
drwxr-xr-x 11 sybase sybase 512 Jan  8 16:14 ..
...
-rw-r--r-- 1 sybase sybase 288 Jan  8 16:14 style.dft
...
(uid=syuid)_sunbox> mv ./style.dft ./style.dft.orig
(uid=syuid)_sunbox> cp ./style.dft.orig ./style.dft
(uid=syuid)_sunbox> vi ./style.dft
(uid=syuid)_sunbox> diff ./style.dft.orig ./style.dft
11a12
> /filter="universal"
(uid=syuid)_sunbox>
```
Sample XML indexing

Section 4

This section demonstrates inserting XML data into the table.

```sql
isql -S ase1501sunbox -U usa -P w2048 -D xml_db
insert address_tbl values
(''<?xml version="1.0"
encoding="UTF-8"?>
<fulldoc>
<address>
<street>mailstop
101</street>
<city>CONCORD</city>
<state>MA</state>
<country>USA</country>
</address>
<filepath>
/sy/ase1501sunbox/EFTS-15_0/collections
/made_using_mkvdk
/xml_data
/row1_well_formed.xml
</filepath>
</fulldoc>
')
go
(1 row affected)
insert address_tbl values
(''<?xml version="1.0"
encoding="UTF-8"?>
<fulldoc>
<address>
<street>building 6</street>
<city>BALTIMORE</city>
<state>MD</state>
<country>USA</country>
</address>
<filepath>
/sy/ase1501sunbox/EFTS-15_0/collections
/made_using_mkvdk/xml_data/row2_well_formed.xml
</filepath>
</fulldoc>
')
go
(1 row affected)
quit
```
Section 5

This section recreates the text index.

```sql
isql -Sase1501sunbox -Usa -P -w2048 Dxml_db1
sp_create_text_index 'textsvr'
'ti_address_tbl', 'address_tbl', " ", 'xmlcol'
go
(return status = 0)
quit
```

Section 6

This section issues select requests that isolate rows based on the XML data, using the text index.

```sql
isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
t1.id
from
ti_address_tbl t1,
address_tbl 2
where
t1.id = t2.id and
  t1.index_any = 'USA <in> address'
go
quickpass disabled, reason: local table
SELECT
id
FROM
  xml_db.dbo.ti_address_tbl
WHERE
  (index_any = 'USA <in> address')
id
--------------
1
2
(2 rows affected)
quit
```

```sql
isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
```
t1.id
from
ti_address_tbl t1,
    address_tbl t2
where
t1.id=t2.id and
t1.index_any = 'USA <in> country'
go
quickpass disabled, reason: local table
SELECT
    id
FROM
    xml_db.dbo.ti_address_tbl
WHERE
    (t1.index_any = 'USA <in> country')
id
-----------------
  1
  2
(2 rows affected)
quit

isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
t1.id
from
ti_address_tbl t1,
    address_tbl t2
where
t1.id=t2.id and
t1.index_any= 'USA'
go
quickpass disabled, reason: local table
SELECT
    id
FROM
    xml_db.dbo_ti_address_tbl
WHERE
    (t1.index_any = 'USA')
id
---------------
1
2
(2 rows affected))
quit

isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
t1.id
from
ti_address_tbl t1,
    address_tbl t2
where
  (t1.id=t2.id and
t1.index_any='mailstop <in> address <and>
        MA <in> state'
gogo
quickpass disabled, reason: local table
SELECT
  id
FROM
    xml_db.dbo_ti_address_tbl
WHERE
  (index_any ='mailstop <in> address <and> MA
       <in> state'
  )
id
---------------
1
(1 row affected))
quit

isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
t1.id
from
to_address_tbl t1
    address_tbl t2
Sample XML indexing

where

t1.id=t2.id and
t1.index_any =
'sy <in> filepath <and>
asesunbox <in> filepath <and>
EFTS-15_0 <in> filepath
<and> collections <in> filepath <and>
made_using_mkvd <in> filepath <and>
xml_data <in> filepath <and> ro* <in> filepath'
go
quickpass disabled, reason: local table
SELECT
id
FROM
xml_db.dbo.ti_address_tbl
WHERE

( index_any = 'sy <in> filepath <and>
asesunbox <in> filepath <and>
EFTS-15_0 <in> filepath
<and> collections <in> filepath <and>
made_using_mkvd <in> filepath <and>
xml_data <in> filepath <and> ro* <in> filepath'
id )
----------
1
2
(2 rows affected)
quit

isql -Sase1501sunbox -Usa -P -w2048 -Dxml_db
select
t1.id
from
ti_address_tbl t1,
address_tbl 2
where
t1.id = t2.id and
t1.index_any =
'sy <in> filepath <and>
SELECT
  id
FROM
  xml_db.dbo.ti_address_tbl
WHERE
  (index_any = 'sy <in> filepath <and>
  ase1501sunbox <in> filepath <and>
  EFTS-15_0 <in> filepath <and>
  collections <in> filepath <and>
  made_using_mkvdk <in> filepath <and>
  xml_data <in> filepath <and> row2* <in> filepath')

id
---------
2
(1 row affected)
quit
Sample XML indexing
Symbols

, (comma)  in SQL statements  xv
{} (curly braces)  in SQL statements  xv
... (ellipsis) in SQL statements  xvii
() (parentheses)  in SQL statements  xv
[] (square brackets)  in SQL statements  xv
<> (angle brackets), enclosing Verity operators in  51

See square brackets [ ] and angle brackets < >

A
accrue operator  50, 52
Adaptive Server  
  connecting to an Enhanced Full-Text Search engine  1  
  processing a full-text query  10
and operator  50, 53
  with the not modifier  60
angle brackets (< >), enclosing Verity operators in  51
  attention events, tracing  73
  Open Server  74
auto_online configuration parameter  22, 68, 70, 162

B
backDir configuration parameter  68, 70, 77, 157
backup and recovery  75
backup files  
  default location of  68, 70
backup operations, tracing  73
batch_blocksize configuration parameter  67
batch_size configuration parameter  67, 69
  and performance  82–83
brackets.
creating 143
default character set 71
default language 70
disabling optimization 73, 79
displaying the names of 146
dropping 145
location of 6
setting the location of 67
modifying data in 23
optimizing 147
performance issues when updating 83
populating with data 20
and reindexing 148
restoring from backup 75, 77
columns
valid datatypes to index 5
comma (,) in SQL statements xv
commands in Verity.
See operators (commands)
complement operator 50, 53
Component Integration Services
connecting to an Enhanced Full-Text Search engine 85
catalog file
sample 163
catalog parameters 67–68, 69–70
See also individual catalog parameters
auto_online 162
backDir 77, 157
batch_size parameter and performance 82–83
charset 71
cluster_effort 48, 152
cluster_max 48, 152
cluster_order 48, 152
cluster_style 48, 152
displaying values 154
language 70
max_sessions parameter and performance 83
min_sessions parameter and performance 83
modifying values 154
nocase 75
sort_order 46, 72
srv_traceflags 74
vdkCharset 71
vdkLanguage 70
catalog parameters, Adaptive Server

cis cursor rows 81
cis packet size 82
connecting to an Enhanced Full-Text Search engine 85
catalogs, number of user 83
conventions
See also syntax
directory paths xiv
used in manuals xv
curly braces ({}),
in SQL statements xv
cursor events, logging 73
custom thesaurus 34
and creating the control file 35
and examining the default thesaurus 35
and the mksyd utility 37
and replacing the default thesaurus 37

D
databases
bringing online for full-text searches 22
databases, bringing online automatically 68, 70
datatypes
and indexing 21
of indexed columns 5, 143
default_Db configuration parameter 68, 70
defining multiple Enhanced Full-Text Search engines 15
delete operations
creating triggers for 23
deletes
and updating the text indexes 8
from the text_events table 141
from the vesaux table 142
document filters 6
document zones
and multiple columns in a text index 22
using with the in operator 53
dump database command
and the sp_text_dump_database system procedure 77, 157

184 Enhanced Full-Text Search Specialty Data Store
E
ellipses (…) in SQL statements xvii
Enhanced Full-Text Search engine
  changing the name of 15
  configuring multiple engines 15, 84–85
  connecting to 85
  document filters 6
  how queries are processed 10
  notifying of updates to the text_events table 161
  operators 50–60
  relationship of components 9
  shutting down 66
  starting as a service 65
  starting for UNIX platforms 63
  starting for Windows NT 65–66
  starting with Sybase Central 65
Enhanced Full-Text Search Specialty Data Store
  components of 5–9
error log file
  setting the path name of 68
  specifying in the runserver file 64
error logging 73
errorLog configuration parameter 68, 70
events, logging 73–74

F
fields
  formatting XML data, sample of 174
fields, formatting XML data for 173
file descriptors
  setting limits 68
filters, document 6
  creating 32
  and document zones 54
forceplan
  and forcing join orders 81
full-text search queries
  bringing databases online for 22
  and case sensitivity 52
  components of 43
  processing a 10
  and requesting clustered result sets 49
  sort order specifications 46–47
  and using topics 41
  using alternative syntax 52

G
getsend program 169

H
highlight pseudo column 44

I
id pseudo column 7, 45
  mapping to the IDENTITY column in the source table 20
  and query optimization 81
IDENTITY columns
  adding a unique index 20
  adding to a source table 20
  adding to existing source table 20
  displaying with the text index 146
  example of adding 25
  joining with the index table in the source table 7, 10
in operator 50, 53
index table
  contents of 7
  creating 20, 143
  dropping 145
  and the id column 19
  in a query 10
  joining with the source table 7
index_any pseudo column 45
  and query optimization 81
insert operations
  creating triggers for 23
  inserts
    and updating the text indexes 8
install event installation script
  editing 17
  example of using 25
  using 17
installtextserver installation script

Users Guide
Index

and creating multiple Enhanced Full-Text Search engines 84
editing 15
location of 15
instsvr.exe utility 66
integrity, maintaining 6
Intelligent Classifier 40
interfaces
  tracing calls between Enhanced Full-Text Search engine and Verity 73
interfaces configuration parameter 68, 70
interfaces file
  setting the location of 68, 70
  specifying in the runserver file 64

J
join order
  ensuring correct 80
joining the source table with the text index 5, 7, 10, 19, 43
  and increasing performance of 80

K
keys modifier 36
knowledge base map
  creating 40
  defining the location of 41
knowledge_base configuration parameter 41, 68, 70

L
language
  setting the default 70–71
language configuration parameter 67, 69
  setting the default 70
language events, logging 73
like operator 50, 54
  enabling literal text in the QBE specification 27
limits
  file descriptors 68
list
  keyword 36
logging events using trace flags 73–74

M
maintaining integrity 6
many operator modifier 60
max_docs pseudo column 45
  with clustered result sets 49
  and increasing query performance 80
  and sort orders 72
max_indexes configuration parameter 67, 69
max_packetsize configuration parameter 67, 69
max_session_fd 68, 70
max_sessions configuration parameter 67, 69
  and performance 83
max_stacksize configuration parameter 67, 69
max_threads configuration parameter 67, 69
metadata 6
min_sessions configuration parameter 67, 69
  and performance 83
mksyd utility
  and creating a custom thesaurus 37
  and examining the default thesaurus 35
mktopics utility 40
multiple users 85

N
naming the Enhanced Full-Text Search engine 68, 70
near operator 50, 54, 55
near/n operator 50, 55
  with the order modifier 61
network requests, tracing 74
nocase configuration parameter 68, 70, 75
not operator modifier 60

O
online databases.
  See databases, bringing online
Open Server events, tracing 74
Open Server trace flags 74
operator modifiers

Enhanced Full-Text Search Specialty Data Store
Index

case  60
many  60
not  60
order  61

operators (commands)  50–60
accrue  50, 52
and  50, 53
complement  50, 53
enclosing in angle brackets  51
in  50, 53
like  50, 54
near  50, 54, 55
near/n  50, 55
or  50, 53
paragraph  50, 55
phrase  50, 55
product  50, 56
and relevance-ranking  45–46
sentence  50, 56
stem  50, 56
sum  50, 57
thesaurus  50, 57
topic  50, 58
wildcard  50, 58
word  50, 59
yesno  50, 60

optimization, disabling  73, 79
or operator  50, 53
with the not modifier  60
order operator modifier  61

outline file for topics  39

P

paragraph operator  50, 55
with the many modifier  60
with the order modifier  61
parameters
of a search  7
parentheses ()
in SQL statements  xv
performance and tuning
adding a unique index  20
and using multiple Enhanced Full-Text Search engines  84

disabling text index optimization  79
increasing query performance  80–81
reconfiguring Adaptive Server  81–82
reconfiguring the Enhanced Full-Text Search engine  82–83
and sp_text_notify  83
phrase operator  50, 55
with the many modifier  60

procedures.
See system procedures
processed events
removing from the text_events table  141
processing full-text searches  10
product operator  50, 56
propagating changes to the collections  8
proxy tables as a source table  6
pseudo columns  7
cluster_keywords  44, 49
cluster_number  44, 49
highlight  44
id  45
in a query  10
index_any  45
max_docs  45, 49
score  45–46
sort_by  45, 46–47, 49
summary  45, 47

Q

QBE specification.
See query-by-example queries
and pseudo columns  7
queries, full-text search
bringing databases online for  22
and case sensitivity  52
components of  43
ensuring the correct join order  80
increasing performance of  80–81
processing of  10
requesting clustered result sets  49
sort order specifications  46–47
and using topics  41
using alternative syntax  52

Users Guide
Index

query-by-example
  configuring for all tables 29
  configuring for individual tables 29
  enabling 27
  and the like operator 54

R
  ranking documents.
    See relevance-ranking
  recovery 75
    and synchronizing a text index with the source table 148
  relevance-ranking 45–46
    See also score pseudo column
  remote procedure calls
    sp_traceoff 74, 80
    sp_traceon 74, 80
  remote tables as a source table 6
  replicating text indexes 24
  RPC events, logging 73
  RPCs.
    See remote procedure calls
  runserver file 63
    flags for 63

S
  sample files
    configuration file 163
    illustrating clustering 168
    illustrating custom thesaurus 35, 168
    illustrating query-by-example 168
    illustrating summarization 168
    illustrating topics feature 39, 168
  sample of formatting XML data for fields and zones 174
  sample program getsend 169
  sample scripts
    sample_text_main.sql 19, 24, 167
  score pseudo column 8, 45–46
    with clustered result sets 49
    and default sort order 72
    and the many modifier 60
    sorting by 47
    values
    how Sybase reports 46
  scripts, sample
    sample_text_main.sql 19, 24, 167
  search parameters 7
  sentence operator 50, 56
    with the many modifier 60
    with the order modifier 61
  sessions, number of user 83
  showplan
    and examining join orders 81
    shutting down the Enhanced Full-Text Search engine 66
  sort orders
    and clustered result sets 47, 49
    by column 30, 47
    in a query 46–47
    max_docs and sort order 72
    by score 47
    setting the default 72
    by timestamp 47, 72
  sort specifications
    setting up a defined column to sort by 30
  sort_by pseudo column 45
    and requesting a clustered result set 49
    and specifying a sort order 46–47
    and setting up a defined column as a sort specification 30
  sort_order configuration parameter 46, 68, 70, 72
  source tables
    adding an IDENTITY column to 19
    changes to data 149, 161
    contents of 5
    and displaying text indexes 146
    in a query 10
  sp_addserver system procedure 84
  sp_check_text_index system procedure 140
  sp_clean_text_events system procedure 141
  sp_clean_text_indexes system procedure 142
  sp_create_text_index system procedure 20, 143–144
    creating indexes that use a filter 32
    example of using 26
    specifying multiple columns 22
  sp_drop_text_index system procedure 145
  sp_help_text_index system procedure 146
  sp_optimize_text_index system procedure 79, 147

Enhanced Full-Text Search Specialty Data Store
**Index**

- **sp_redo_text_events** system procedure 148
- **sp_refresh_text_index** system procedure 149–150
  - modifying data in the collections 23
  - running automatically 23
- **sp_show_text_online** system procedure 151
- **sp_statistics** system procedure
  - disabling 73, 80
- **sp_text_cluster** system procedure 152–153
- **sp_text_configure** system procedure 69, 154–155
- **sp_text_dump_database** system procedure 76, 156–158
- **sp_text_kill** system procedure 159
- **sp_text_load_index** system procedure 77, 159–160
- **sp_text_notify** system procedure 161
  - and modifying data in the collections 23
  - and performance issues 83
  - and turning off optimization 79
- **sp_text_online** system procedure 23, 162
  - example 26
- **sp_traceoff** remote procedure call 74, 80
- **sp_traceon** remote procedure call 74, 80
- SQL parsing, tracing 73
- square brackets [ ] in SQL statements xv
- **srv_traceflags** configuration parameter 68, 70, 74
- starting the Enhanced Full-Text Search engine
  - from Sybase Central 65
  - on UNIX platforms 63
  - on Windows NT 65–66
  - as a service 65
- **startserver** utility 63
- start-up
  - and setting the number of user connections 83
  - start-up commands
    - and the runserver file 63
    - on Windows NT 65
- **stem** operator 50, 56
  - with the **many** modifier 60
- **style.dft** file 32
- **style.prm** file
  - editing an existing collection’s 144
  - editing for an existing collection 29
  - editing the master 28
  - and enabling Verity functionality 27
  - location of an existing collection 29
  - location of master 29
- **style.aff** file 30, 32
- **style.vgw** file 30, 32
- **sum** operator 50, 57
  - summarization
    - configuring for all tables 29
    - configuring for individual tables 29
    - enabling 27
    - writing queries requesting 47
  - **summary** pseudo column 45
    - enabling before using 27
    - using 47
  - Sybase Central, starting from 65
  - symbols in SQL statements xv
  - synonym list for a custom thesaurus 35
- **synonyms**
  - statement 36
  - syntax conventions, Transact-SQL xiv
  - syntax, alternative Verity 52
  - **sysservers** table
    - adding Enhanced Full-Text Search engines 84
- system procedures
  - See also individual system procedures
  - **sp_check_text_index** 140
  - **sp_clean_text_events** 141
  - **sp_clean_text_indexes** 142
  - **sp_create_text_index** 143–144
  - **sp_drop_text_index** 145
  - **sp_help_text_index** 146
  - **sp_optimize_text_index** 147
  - **sp_redo_text_events** 148
  - **sp_refresh_text_index** 149–150
  - **sp_show_text_online** 151
  - **sp_text_cluster** 152–153
  - **sp_text_configure** 154–155
  - **sp_text_dump_database** 156–158
  - **sp_text_kill** 159
  - **sp_text_load_index** 159–160
  - **sp_text_notify** 161
  - **sp_text_online** 162
- system tables
  - updating 139

**T**

- TDS data, tracing 74
Index

TDS headers, tracing 74
TDS tokens, tracing 74
text documents, types of 6
text indexes 76, 77
   backing up 156
   bringing online 162
   creating 20, 143
   creating and batch sizes 82
   displaying a list of 146
   displaying online 151
   dropping 145
   example of creating 24–26
   and the index table 7
   metadata 6
   that include multiple columns 22
   optimizing 147
   performance issues when updating 83
   placing on multiple Enhanced Full-Text Search engines 84
   and reindexing 148
   replicating 24
   restoring from backup in Enhanced version 75
   setting location of backup files 68, 70
   and tracing information 73
   update using text_events table 8
   updating 79
   using a document filter with 32

text_db database 6, 75, 76
   backing up 156
   changing the name of 15, 18
   restoring from backup 77
   and the vesaux table 6
   and the vesauxcol table 7
text_events table 8, 76
   changing the status of entries 148
   columns in 8
   creating 17
   example of creating 25
   recording inserts, updates, and deletes 149
   removing entries from 141
   restoring from backup 76, 77
   and sp_text_dump_database 76, 157
   and sp_text_load_index 78
textsvr.cfg file
   sample 163

thesaurus operator 50, 57
using a custom thesaurus 34
thesaurus, custom 34
   and creating the control file 35
   and examining the default thesaurus 35
   and the mksyd utility 37
   and replacing the default thesaurus 37
timestamp
   sorting by 72
topic operator 41, 50, 58
topic set directories 40
   mapping to 40
topics
   creating a knowledge base map 40
   creating a topic set directory 40
   creating an outline file 39
   creating complex relationships 39
   description of 38
   executing queries using 41
   sample files 39
   troubleshooting 42
trace flags 73
   enabling trace flags 11 and 12 79
   Open Server 74
   setting to examine join orders 81
traceflags configuration parameter 68, 70
triggers for running sp_refresh_text_index 23

U
Unicode 171
   and wildcard searches 73
Unicode support 171
unique index
   adding to an IDENTITY column 20
   example of creating 25
update operations
   creating triggers for 23
update statistics
   disabling 80
updates
   and updating the text indexes 8
   updating indexes 79
user
   connections 83
   sessions 83
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>V</strong></td>
</tr>
</tbody>
</table>
| **vdkCharset** configuration parameter | 67, 69  
| setting the default | 71  
| **vdkHome** configuration parameter | 67, 70  
| **vdkLanguage** configuration parameter | 67, 70  
| setting the default | 70  
| **Verity** |  
| setting the Verity directory | 67  
| tracing Verity processing | 73  
| **Verity collections** |  
| **Verity query** |  
| **vesaux** table |  
| columns in | 6  
| creating entries | 144  
| removing entries from | 142  
| removing entries when dropping text indexes | 145  
| updating | 20  
| **vesauxcol** table |  
| columns in | 7  
| removing entries when dropping text indexes | 145  
| updating | 20  

**W**  

**wildcard** operator | 50, 58  
| using with data in Unicode format | 73  
| with the **case** modifier | 60  
| with the **many** modifier | 60  
| **Windows NT** |  
| directory paths | xiv  
| **word** operator | 50, 59  

**X**  

XML data, formatting for fields and zones | 173  
XML data, formatting for fields and zones, sample of | 174  

**Y**  

**yesno** operator | 50, 60  

**Z**  

zones.  
**See** document zones

user databases | 75, 76  
backing up | 156  
bringing online automatically | 68, 70  
bringing online for full-text searches | 22, 162  
displaying a list of text indexes for | 146  
displaying online | 151  
restoring from backup | 77  
user table.  
**See** source table

with the **case** modifier | 60  
with the **many** modifier | 60  
**writetext** command, using triggers with | 23  

**users Guide**  

191