IBM

IBM Text Analyzer Business Component

Installation Guide

Version 1.1

… A member of the WebSphere Business Components family

Before using this information and the product it supports, be sure to read the general information under "Notices" on page 8.

First Edition (November 2000)

This edition applies to version 1.1 of IBM Text Analyzer Business Component (part number 20P4393 or 20P4395), and to all subsequent releases and modifications until otherwise indicated in new editions. Make sure you are using the correct edition for the level of the product.

Corrections and suggestions for future revisions of this document are appreciated. Mail your comments to:

IBM Canada Ltd. Laboratory
Information Development
2G/KB7/1150/TOR
1150 Eglinton Avenue East
Toronto, Ontario, M3C 1H7
Canada

When you send information to IBM, you grant to IBM a nonexclusive right to use or distribute the information in any way they believe appropriate without incurring any obligation to you.

Contents

IBM Text Analyzer Business Component Version 1.1 Installation Guide

This installation guide contains a section with installation instructions for Text Analyzer.

The green text represents changes that were made to this document after the Text Analyzer Business Component Version 1.1 CD was created.

Text Analyzer installation instructions

The IBM Text Analyzer (TA) Business Component is installed as two distinct subcomponents:

- Categorization Engine (CE)

- Training Tool (TT)

The Categorization Engine is installed and deployed as a WSBC Version 1.1 Advanced Component. The Training Tool is installed and deployed as a Java API. The Text Analyzer setup which runs from the IBM Text Analyzer Business Component Version 1.1 CD-ROM will copy all required Text Analyzer files to your WSBC Studio installation.

Development installation prerequisites

The following are the prerequisites for installing TA files to a development workstation from the CD-ROM.

Software

The Text Analyzer installation procedure has a dependency on WebSphere Business Components Studio. You must install WSBC Studio to your development workstation before you will be able to install any Text Analyzer component files to the workstation. You should follow the recommended order of installation for any software you expect to use; this will facilitate automatic setup (See the WSBC Studio installation guide).

The Text Analyzer documentation installation procedure has a dependency on the WSBC Studio documentation system. Both the WSBC Studio documentation system and the TA documentation may be installed separately from other WSBC installable components by selecting custom installation options.

Hardware

The Text Analyzer installation procedure does not have any hardware prerequisites for installation beyond those required by WSBC Studio. A full installation of Text Analyzer, including documentation, requires 73539Kb of available hard drive space.

Development deployment prerequisites

The Text Analyzer AC environment must provide the Java API for XML Parsing Version 1.0.1 (At the time of this writing this version is being replaced by Java API for XML Processing Version 1.1). This is available by downloading it as set of Java packages from the Sun Java Web site and is deployed as the jaxp.jar and parser.jar files. This API is used to access the XML4J XML Parser which is deployed as part of WebSphere Application Server.

Production deployment prerequisites

The Text Analyzer AC has a runtime dependency on IBM XML4J XML Parser for Java and the Java API for XML Parsing.

If you have WebSphere Application Server installed in your production environment, then you already have the required xml4j.jar file. You may obtain the parser separately from the IBM alphaWorks Web site. When deploying the Categorization Engine or Training Tool you may also need to deploy XML4J by copying the .jar file to the deployment environment, specifying the path to the XML4J .jar file or both. See "Development deployment prerequisites" for the details of how to obtain and deploy the Java API for XML Parsing.

Install Text Analyzer files to a development workstation

Text Analyzer files are installed to a Windows NT development workstation by running the setup program from the IBM Text Analyzer Business Component CD-ROM. This file is located in the TAInstall folder and is named Setup.exe.

Run the Setup.exe file to start the Text Analyzer installer and follow the prompts. You will be warned if you have not installed WebSphere Application Server and Java API for XML Parsing (JAXP); you may continue with the installation and install these items at a later date.

You must accept the International Program License Agreement to start the installation.

Select the installation components you wish to install. Choose **Typical** to install both the Text Analyzer files and the Text Analyzer documentation. Choose **Custom** to select individual installation components (uncheck the components that should not be installed.)

If you selected installation of the Text Analyzer component itself, then the Text Analyzer files will be copied to a TextAnalyzer folder within your WSBC Studio installation. If you selected to install the Text Analyzer documentation, the WSBC online help system will be updated; you will find the Text Analyzer documentation as an item under the Topics heading in the table of contents.

A Text Analyzer folder is created at *<Your WSBC Install directory>*\1.1\ACFeatures\Components\TextAnalyzer\. All Text Analyzer files for both the Text Analyzer AC (Categorization Engine) and the Training Tool are located here. You will need to copy various files from this location when developing a TA client, developing a customized TT, and when installing and deploying the CE and TT to a production environment.

Installed files

The following tables describe the Text Analyzer files and their locations after installation. The default installation folder `C:\Program Files\IBM\WSBC\1.1\ACFeatures\Components\` contains the TextAnalyzer folder; all Text Analyzer files are located within this folder. The xml4j.jar (XML4J) and parser.jar (JAXP) files are required to complete an installation as described in the prerequisite sections.

Text Analyzer folder	
File name	**Description**
Events.xsd	TextAnalyzer AC event data definition schema
Functions.xsd	TextAnalyzer AC functions definition schema
TextAnalyzerAC.jar	Text Analyzer AC implementation
TextAnalyzerACEJB.jar	Text Analyzer deployable EJB implementation
TextAnalyzerACEJBClient.jar	Interface stubs for remote Text Analyzer clients

Text Analyzer folder	
File name	**Description**
TextAnalyzerACEJBDeployed.jar	WebSphere deployed Text Analyzer AC
TextAnalyzerBase.jar	Training Tool and Categorization Engine

TextAnalyzer\is	
File name	**Description**
InterfaceModels.xsd	WSBC TextAnalyzerAC type definition schema

TextAnalyzer\LPResources\AIX\bin	
File name	**Description**
libeflnl27x.so	AIX language processing executable
libPoeJNI.so	AIX JNI for language processing

TextAnalyzer\LPResources\Solaris\bin	
File name	**Description**
libeflnl27s.so	Solaris language processing executable
libPoeJNI.so	Solaris JNI for language processing

TextAnalyzer\LPResources\WIN\bin	
File name	**Description**
eflnl27w.dll	Windows language processing executable
poejni.dll	Window language processing JNI

TextAnalyzer\LPResources\dictionaries	
File name	**Description**
files named *.dic and *.abr	Language specific dictionaries and abbreviations. See online Text Analyzer references for the complete list of supported languages.

TextAnalyzer\rose\TextAnalyzerAC\Interface	
File name	**Description**
TextAnalyzerAC.cat	Model of the Text Analyzer AC and its interface

TextAnalyzer\TTGUIV1R1	
File name	**Description**
data.xml	Training sample template
projectini.xml	Project configuration template
projectList.xml	Project list
TTGUI.bat	V1R1 GUI launcher

TextAnalyzer\TTGUIV1R1\Samples\FullReuters-21578	
File name	**Description**
data.xml	Training sample
projectini.xml	Project configuration
rules.xml	Categorization rule set

TextAnalyzer\TTGUIV1R1\Samples\MiniReuters-21578	
File name	**Description**
data.xml	Training sample
projectini.xml	Project configuration
rules.xml	Categorization rule set

Install the TT Java packages to a VisualAge for Java workspace

1. Create a VAJ project in your workspace for the TA objects.

2. Import all .class and resource files from the TABase.jar file into your selected project.

Install the TA AC (CE) Java packages to a VisualAge for Java workspace

1. Click on the EJB tab in the VAJ workbench.

2. Create an EJB group if required. Right-click on the Enterprise Bean pane to open the context menu and select **Add** from the menu and then select **EJB Group** from the submenu.

3. Right-click on the EJB group to open the context menu and select **Import Enterprise Bean...** to open a dialog box.

4. Browse to the TextAnalyzerAC.jar . Ensure that all Beans, classes, and resources are selected with check marks.

5. Click **Finish** to load the AC.

6. Ensure that the CommandTarget Session Bean is available from the AC Services package located in an EJB group. If it is not installed, install it from CommandTargetEJB.jar.

Deploy the TT API in a development environment.

Install the following files to your Training Tool test environment:

- TextAnalyzerBase.jar

- eflnl27w.dll

- POEJNI.dll

- Language dictionaries as required (*.dic and *.abr files)

Refer to the online documentation for details on how to deploy and configure the V1R1 Training Tool reference GUI.

Deploy the TA AC EJB in the VAJ WebSphere test environment

1. Open the workspace EJB Group context menu with a right-click. Select **Generate Deployed Code** from the menu to generate the EJB code.

2. Again, open the workspace EJB Group context menu with a right-click. Select **Add** from the menu and then select **Server Configuration** from the submenu.

3. Start the WebSphere Test Environment from **Tools** > **WebSphere Test Environment** and in the left pane select Servlet Engine. In the right pane click on **Edit Class Path**, click the **Select all** button and click **OK**. Click the **Start Servlet Engine** button.

4. Start the Persistent Name Server by selecting the **Persistent Name object** and then in the right pane selecting **Start name server**. Check the console and wait for the "Server open for business" message.

5. Start the EJB Server. From the Workbench select the EJB tab, right-click in the Enterprise Bean pane, and select **Open to** > **Server Configuration**. You should see the ACCommand and TextAnalyzer servers. Right-click on each server and select Start; wait for the "Server open for e-business" message to appear in the console window.

Deploy the TT Java packages in a production environment.

Install the following files to your Training Tool production environment:

- TextAnalyzerBase.jar
- eflnl27w.dll
- POEJNI.dll
- Language dictionaries as required (*.dic and *.abr files)

Install a deployable TA AC EJB on a production WebSphere Application Server

Install the following files to your production environment.

- TextAnalyzerBase.jar
- TextAnalyzerAC.jar
- TextAnalyzerACEJBDeployed.jar
- XML4J.jar
- efln27w.dll
- POEJNI.dll
- Language dictionaries as required (*.dic and *.abr files)

Deploy a TA AC instance using a production WebSphere Application Server

Before deploying Text Analyzer you should deploy AC services. (See "Deploy Advanced Component Services on WebSphere Application Server" in the WSBC Studio Installation Guide.)

Configure WebSphere Application Server to deploy Text Analyzer AC instances

Before proceeding, ensure that the WebSphere Application Server is configured for use with WSBC ACs. The ACServicesEJB.jar file should be in the WebSphere Application Server admin classpath. (The TextAnalyzer AC is deployed as an Enterprise Java Bean (EJB)).

1. Start the WebSphere Application Server AdminServer service.

2. Open the WebSphere Administrator's Console.

3. Select the topology view of and navigate to and select the server on which you want to deploy the Text Analyzer AC. A pane will appear on the right side.

4. Enter the following command line arguments into the Command line argument text field on the general tab of the configuration.

   > -mx128m -classpath <*your full path*>/TextAnalyzerBase.jar; <*your full path*>/TextAnalyzerAC.jar;

   Be sure to specify the full file system paths for the .jar files in the classpath argument and to separate the entries with a semicolon. The memory allocation option may be adjusted as required to suit your needs.

Deploy an instance of the Text Analyzer AC using WebSphere Application Server

1. Start WebSphere Application Server service.

2. Start the WebSphere Application Server Administrator's Console.

3. Select the topology view.

4. Navigate down the server tree down to your EJB container.

5. Open the Create EJB dialog with a right-click on the container to open the context menu, select **Create** from the menu, and finally select **Enterprise Bean** from the submenu.

6. Click on the **Browse** button and navigate to the location of the TextAnalyzerACEJB.jar file and select the file to view the EJBs it contains.

7. Select the com.ibm.wsbc.ac.textAnalyzerInterface/TextAnalyzerInterface.ser EJB. If you are prompted to enable work load management, select **No**. (The navigator closes and you are returned to the Create EJB dialog which now has fields filled in.)

8. Ensure that the EJB has a unique JNDI HomeName within the container. Click **Edit** on the Create EJB dialog to open the Deployment Properties dialog. Edit the **JNDI HomeName field**; the default name is the fully qualified package name followed by the Session Bean identifier followed by the instance name (com/ibm/wsbc/ac/textanalyzer/TextAnalyzerInterface). Any name may be used provided that it is unique. Click **OK** to close the Deployment Properties Dialog.

9. Click **Create** and navigate to the location of the TextAnalyzerACEJBDeployed.jar. Select the file to create the Text Analyzer AC instance. Click **OK** to close the Create EJB dialog.

10. Run the AC Deployer Tool to perform the JNDI registration.

Notices

IBM may not offer the products, services, or features discussed in this document in all countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Canada Ltd.,
Department 071,
1150 Eglinton Avenue East
Toronto, Ontario, M3C 1H7
Canada

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems.

Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks and service marks

The following terms are trademarks of International Business Machines Corporation in the United States, or other countries, or both:

AIX
CICS
DB2
DB2 Universal Database
e-business
IBM
LANDP
MQSeries
OS/2 Warp
OS/390
RS/6000
SanFrancisco
VisualAge
Visual Banker
WebSphere

Lotus, Domino, Lotus Notes, and Notes Mail are trademarks of the Lotus Development Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

MMX, Pentium, and ProShare are trademarks or registered trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark in the United States, other countries, or both and is licensed exclusively through X/Open Company Limited.

Rational Rose is a registered trademark of Rational Software Corporation.

Other company, product, and service names may be trademarks or service marks of others.

End of document